Contents

Acknowledgements

The origins of this thesis go back quite some time, so there have been a considerable number of people with whom I discussed the issues raised here. I would like to express my gratitude to these people and also want to apologize to everybody whom I may have forgotten to mention below.

The very first foundations of this work have been laid in extensive discussions with John Nerbonne and Carl Pollard, whose share in this thesis goes way beyond some footnotes on "personal communication". At about the same time, I also enjoyed the frequent meetings and discussions with Sergio Balari, who helped me shape and sharpen my ideas about HPSG.

Stephan Oepen deserves a special mention, not only for always patiently listening to me without showing obvious signs of boredom, but also for playing the role of the devil's advocate relentlessly questioning my positions.

From my colleagues at DFKI, Rolf Backofen, Bernd Kiefer, Hans-Ulrich Krieger, Günter Neumann, Ulrich Schäfer, and Stephen Spackman, I learned a lot about formalisms, as well as about many other aspects of natural language processing. I am also grateful to them for providing me with a sophisticated grammar engineering environment, which made it possible for me to go beyond pure paper work and implement my ideas in an advanced NLP system. Walter Kasper was of support in so far as he developed the semantic component to the grammar fragment.

Among my colleagues at the Computational Linguistics Department of Saarbrücken University, I would like to thank above all Tanja Agustinova, Gregor Erbach, Reinhard Karger, Brigitte Krenn, and Karel Oliva for providing an inspiring environment for linguistic discussions. Manfred Pinkal, the second supervisor of this thesis, has been quite critical and also equally helpful in the final stages of the thesis.

Hubert Haider, Tilman Höhle, and Marga Reis do not have a direct connection to this work, but from the close interaction with them during my time at the Institut für Maschinelle Sprachverarbeitung (IMS) in Stuttgart I drew a lot of my knowledge about German Syntax that made its way into this work.

With Robert Borsley, Richard Cooper, Richard Hudson, Tibor Kiss, Erhard Hinrichs, Tsuneko Nakazawa and Mark Steedman I had useful discussions about various subjects brought up in this thesis. Dan Flickinger and Ivan Sag, I want to thank for many lively conversations, for listening and talking to me, respectively.

One of the most fruitful and enlightening cooperations in recent time I owe to K. Vijay-Shanker, from whom I learned tremendously much in many different respects. In this context, I would also like to thank Bob Kasper and the people related to the TAG community, above all Anne Abeillé, Aravind Joshi, Tony Kroch, Owen Rambow, and B. Srinivas, who helped me a lot to broaden my linguistic and formal horizon.

Last but not least, Hans Uszkoreit has been more than just a supervisor to this thesis. At this point, I will not say more than simply that I am very grateful for the support that he gave me.

Saarbrücken, May 1996

Postscriptum

This printed version of the thesis has not been modified in comparison to the 1996 version. Although there clearly has been a host of developments in the meantime which would have been relevant for this topic or which were brought to my attention at a later state, I left it in its original state, as has been the policy for this series of publications, and only corrected typos and grammatical errors. Stefan Müller, when going through the script, not only did a very thorough job in pointing out such errors, but also raised a number of highly interesting questions, which, if they had been answered in a fully satisfatory way, certainly would have greatly improved this thesis.

Saarbrücken, June 1998

Chapter 1

Overview and Background

1.1 Introduction

1.1.1 Objectives and Motivation

Over the last years Head Driven Phrase Structure Grammar (HPSG) has become one of the predominant grammar frameworks in the areas of formal and computational linguistics. Based on precise formal foundations it has been appealing both to computationally oriented researchers who are looking for an expressive formalism which mirrors the state of the art, as well as to formal linguists who wanted a flexible theoretical framework to experiment with and to express linguistic generalizations. As a result, work in HPSG has led both to a series of implementations of grammar fragments, as well as to formalized linguistic descriptions of a large number of different languages.

Our objective here is to contribute to these results from both points of view. The work that is being described below stems from an implementation of a comparatively large fragment of German language, which has been developed with HPSG as the formal and linguistic starting point. It should not come as a surprise that in developing and implementing the fragment, we often deviated from standard assumptions in HPSG. Firstly, it is clearly inevitable that given the size of the framework not all aspects could be worked out in the same degree of detail. Even if the authors adopted a high degree of formal rigour, inconsistencies and incompletely fleshed out definitions were hard to avoid. Secondly, what may appear as an elegant formulation on paper can easily turn into a rather awkward and complicated representation when translated into an implemented formalism. Thus, when it comes to implementations, elegance is not just in the eye of the beholder but may have a direct counterpart in simplicity and efficiency. Last but not least, like many theoretical frameworks HPSG grew out of a particular linguistic environment with one specific language providing the predominant empirical basis. As a consequence, what appears to be a straightforward phenomenon with an apparently simple solution in one language, can easily correlate to a very complex problem requiring different and possibly more sophisticated solutions in another language. Thus, adapting a framework to another language will often change the perspective, under which an issue is addressed, quite radically.

However, even if we attribute a comparatively high value to implementational issues and the language specific phenomena of German, it should be clear that we are not propagating unscrupulous grammar hackery disguised as the inevitable result of the empirical evidence of German. Whenever we propose changes to the grammar framework or the treatment of certain phenomena below, our aim is not only to motivate them by practical considerations and empirical data, but also to achieve a simplification of the theory. Changing the theory just for the purpose of being "innovative" does not lead to any progress. In some cases, we will therefore discuss and explore a larger space of different possible options than is actually necessary to reach a stringent solution. Our motivation here is mainly to allow the reader to judge by himself whether among a range of possible changes the adopted solutions are indeed the optimal and most adequate ones.

In short, what we hope to achieve in this work is to present a novel and coherent description of a fragment of German based on the framework of HPSG, while covering the full range from the formal foundations, through the underlying grammar model, to the description of linguistic data and their representation in a precise formal language.

1.1.2 Main Results

The main results of the thesis concern the following items

- Description of a typed feature formalism ($\mathcal{TDL}$) and suggested extension by *cross-classification* for the treatment of HPSG principles
- Discussion and clarification of central assumptions of HPSG
  - notion of *head* and its defining properties
  - relation between universal and language particular grammars
  - formalization and integration of principles in type hierarchies
- Alternative conception of HPSG phrase structure schemata with
  - reduction and generalization of rule schemata
  - parametrization for different types of languages
  - redefinition of LP constraints for generalized schemata
- Novel approach to functional categories as heads involving
  - definition of functional completeness
  - split of head features
  - specific complementation relation between functional and substantive categories
  - definition of intermediate projections for adjunction
- Analysis of German nominal phrases involving
  - determiners and determinerless constructions

  - attributive adjunction
  - agreement and declension phenomena
  - empty noun heads
  - numerals

- Analysis of finite verb position and clause structure in German involving
  - treatment of word order phenomena
  - complementizers
  - V-final and V-initial clauses
  - separable prefixes
  - relative clauses

While this is just a brief overview of the topics addressed, we give a more extensive summary of the results at the end of this chapter in section 1.3.5. This summary will relate the individual contributions of the thesis to a model of layers of abstraction in computational linguistics described below.

1.1.3 Content

This work is divided into five chapters. In the remainder of this chapter we try to set a general background for the framework of HPSG, by briefly sketching the family of so-called constraint-based formalisms and describing the historical development which eventually led into this grammar framework. We also discuss the architecture and layers of abstractions under which a computational linguistics framework can be viewed and relate the contributions of this thesis to this layered model.

Chapter 2 introduces the specification language $\mathcal{TDL}$ in which our grammar fragment is described and implemented. This chapter is divided into a section on the underlying feature language and a section on the type logics and the different opportunities it provides. We also sketch an extension to the type logics which could be useful for the representation of principles in a type lattice. For readers familiar with typed feature logics, this chapter will bear little new information but could nevertheless help to understand what perspective we are taking.

Chapter 3 contains a discussion of the general framework of HPSG. We address issues such as the relation between language universal and language particular grammar, and discuss the notion of head as one of the central concepts of HPSG.

Based on this general background we propose a different setup for rules and principles in HPSG, which will allow to capture universal generalizations while accounting also for language specific parametrization.

Chapter 4 and 5 contain the main bulk of the grammar fragment for German, as well as further details of the proposed changes to the general design of the theory of HPSG. The first of these chapters deals mainly with the problem of functional categories from the perspective of nominal phrases. However, it also contains proposed solutions to problems related to issues such as adjectival adjunction, nominal ellipsis, declension classes and numerals.

Chapter 5 applies the analysis of functional heads to categories in the verbal domain. It describes how the relationship between complementizers and finite verb position can be captured on this basis and develops an approach which allows to model word order phenomena by a homogeneous structural description.

1.1.4 Implementation and Project Note

With the exception of some minor extensions, the fragment described in the chapters 4 and 5 has been implemented and is embedded in a larger fragment of German, which by and large describes all of the standard constructions of German. Besides the described phenomena, this grammar covers among others the different types of complementation and of adverbial constructions, all major types of sentence mood, a broad range of word order phenomena including different forms of verb cluster formation (involving, for example, modals, temporal auxiliaries, passive constructions), as well as some specialized subgrammars for time and date expressions. Some more recent versions of the grammar build on a revised treatment of subcategorization, allowing a simplified treatment of optionality, and a lexicalized treatment of unbounded dependencies which dispenses with empty terminals for SLASH-termination.

The architecture of the grammar comprises run-time executable lexical rules, multi-word lexemes, a typed interface to morphology, and an interface to a speech act recognition component. The grammar consists of appr. 2,000 type definitions, defining a set of about 20 different instances of rule schemata and 1,000 lexical instances, which consist mainly of closed class items (prepositions, determiners, and other functional categories) and exemplary classes of open class items (nouns, adjectives and verbs). Through default entries a lexical data base of about 120,000 open class items can be accessed.

The grammar has been implemented in the context of the DFKI projects DISCO, "A Dialogue System for Autonomous Cooperating Agents", and PARADICE, "Parametrizable Discourse Core Engine".[1] Both projects were funded by research grants of the German Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF), FKZ ITW 9002 0 and FKZ ITW 9403 respectively. The implementation of the grammar has also been underlying the natural language front end of the COSMA ("Cooperative Schedule Management Agent") system and project (BMBF, FKZ ITW 3030). The grammar as well some of the theoretical concepts developed here served as the basis for the development of an algorithm for compiling HPSG into Tree Adjoining Grammars, which was partly carried out in the VERBMOBIL project (BMBF, FKZ 01 IV 101 K/1).[2] The grammar provided also the linguistic basis for an approach of Explanation-Based Learning (EBL) which was designed for efficient processing of constraint-based grammars.[3]

[1]For a brief description of the architecture of the grammar see [Netter 1993]. For an overview to the entire DISCO system see [Uszkoreit et al. 1994].

[2]See [Netter et al. 1994] and [Kasper et al. 1995].

[3]Cf. [Neumann 1994].

1.2 Constraint-Based Theories

In the following we try to set the general background by giving a brief overview on the family of constraint-based frameworks, both from the point of view of classifying these frameworks according to various parameters and from a historical perspective. The main purpose of this overview is both to highlight some distinguishing properties of HPSG and to illustrate some historical developments which appear to have influenced the basic concepts and shape of HPSG.

1.2.1 Perspectives on Constraint-Based Theories

Theories which can be classified as feature-based, unification-based or generally constraint-based have very much dominated the past decades in computational linguistics. What these theories have in common is a declarative approach to grammar specifications, meaning that the underlying logics can be typically defined without any reference to how these specifications have to be processed. They also share the assumption that all or at least a substantial part of linguistic information can be described by means of attribute value pairs. However, these theories also differ in many crucial ways, both in terms of the formalism and in terms of linguistic assumptions.

Without any claim to completeness, we therefore illustrate in figure 1.1 not only how these frameworks fit onto a historical time axis, but also some of the different aspects under which individual constraint-based theories can be grouped together.

One very central division between the frameworks is given by the two vertical columns in which they occur. While the frameworks on the left have been developed mainly as pure grammar *formalisms*, the frameworks on the right hand side not only introduced a new type of formalism but also came with an explicit and dedicated *linguistic theory*. This does not necessarily mean that the formal frameworks on the right-hand side are bound to a specific linguistic theory, but rather that a primary vehicle for dissemination in these cases appears to have been the linguistic descriptions that came with these formalisms. In contrast, developers of the more theory-neutral frameworks on the left-hand side have stressed that their formalisms and implementations should in principle allow the implementation of *any* linguistic theory and therefore designed them them as open and unrestricted as possible.

However, these theories can also be classified according to the formal concepts they involved. Thus, from another perspective, which cuts across this distinction, frameworks such as GPSG, LFG, PATR-II and in some sense also DCG belong to the same class, as they do not limit the descriptive devices to pure feature

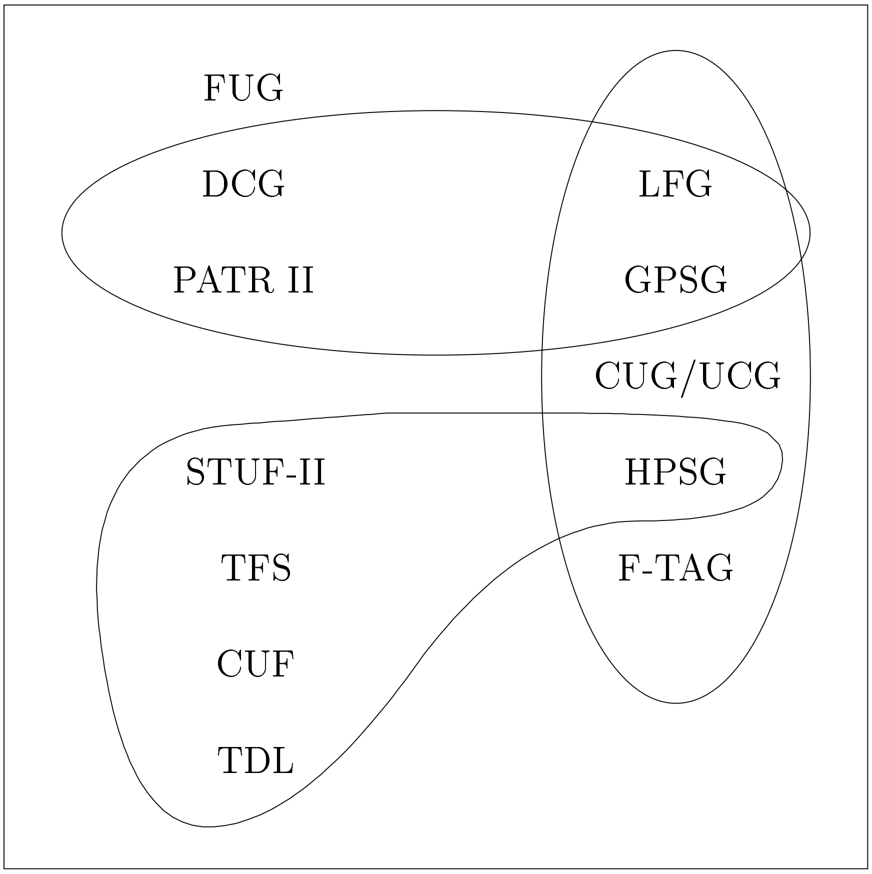


Figure 1.1: Constraint-based Frameworks in Computational Linguistics

constraints, but also employ a context-free backbone as a central device for describing constituent structure. Although feature-based TAGs could be also seen as part of this group, they go beyond simple context-free backbones, as they do not constrain the description of constituent structure to trees of depth one.

A third classification groups together those frameworks that on the time axis occur below UCG. All of these theories and formalisms, with the notable exception of F-TAG, employ types or sorts in addition to pure feature constraints. It is interesting to see that none of these theories retained context-free backbones as an explicit descriptive device, even if some of them build on the derivation of such backbones for processing.

Of course, classifications of this kind would not be very useful, if they didn't also help to understand the historical development in this area of computational linguistics. In the next section we try to make these tendencies clearer by briefly looking at the individual contributions that these frameworks made.

1.2.2 Historical Overview

Functional Unification Grammar – FUG It should not come as a surprise that Martin Kay's Functional Unification Grammar [Kay 1979] occurs at the root of all the frameworks. It appears that this formalism was among the first to emphasize notions such as declarativity, monotonicity and underspecification by making crucial use of feature unification as a means of processing grammatical information. From today's perspective, the fact that this formalism never became really wide-spread as an implementational basis for grammar development could be only explained through the lack of a comprehensive linguistic theory associated with the framework. To appreciate the contributions of FUG as an unrestricted specification language, it has to be kept in mind that at that time generative transformational grammar was still in its prime, and the scientific discussions concentrated on how the constraints on linguistic formalism can mirror the constraints imposed by linguistic theory. In contrast, FUG was developed as a (non-transformational) grammar formalism which did not impose any restrictions on the theory of grammar. However, this was not the only sense in which FUG appears to have been almost too far ahead of its time. FUG already involved a separation of immediate dominance and linear precedence relations, as it was later suggested in GPSG, and it anticipated the description and phrase structure constituents by means of feature structures as well as its close integration with other levels of abstraction, as we find it in CUG and HPSG,

Definite Clause Grammar – DCG Definite Clause Grammars, developed by [Pereira and Warren 1980], represent a formalism, which after its invention has often been used as the basis for efficient implementations of other grammar formalisms. In contrast to later unification-based formalisms, DCG presupposes a fixed arity of predicates, i.e., a fixed number and order of features associated with a node. Since this format only requires term unification rather than graph unification, DCG can be directly implemented in programming languages such as PROLOG, providing this kind of resolution as part of the programming language.

Lexicalized Functional Grammar – LFG Almost in parallel to FUG the framework of Lexical Functional Grammar (cf., for example [Bresnan 1982b], [Kaplan and Bresnan 1982]) was developed, which aimed to provide a synthesis of three different components; a precise formal description language, a comprehensive description of a whole range of linguistic phenomena and a realistic model of processing. The strong emphasis on the latter aspect can also be understood as a reaction to transformational models, which at that time had been proven to be computationally untractable, unless they obeyed certain constraints [Peters and Ritchie 1973].

On the formal side and from the point of view of processing, LFG claimed to have achieved the ideal compromise between expressivity and efficiency by combining an explicit context-free backbone with constraint annotations to phrase structure rules. An important notion that was brought up in this context was the distinction between constraint descriptions (functional descriptions) and models satisfying these constraints. The only operation employed in the constraints was equality, such that derivations could make use of simple algorithms for equation solving.[4]

At least in its earlier days, the linguistic claims of LFG also have to be seen as a strong reaction against the transformational tradition.[5] It showed that all structural transformations can be abandoned, if one postulates rules for transformations of lexical items. Since these lexical rules could (or in fact had to) be applied at compile time, run-time decidability was given, if certain constraints (off-line parsibility) were obeyed. Secondly, LFG rejected the view that grammatical relations have to be encoded configurationally by means of phrase structure positions and introduced sets of grammatical function labels as primitives of the theory.

Generalized Phrase Structure Grammar – GPSG Like LFG, Generalized Phrase Structure Grammar [Gazdar et al. 1985] came with a substantial grammar fragment of English, which was specified in a formally precise description language. GPSG's most important contribution (but maybe also its most strangling claim) was the attempt to show that structural transformations not only could be abandoned, but also that the computational complexity of grammars for natural languages could be restricted to the complexity of context-free languages. While this possibility had been debated before, it had been rejected on the basis of the argument that pure context-free phrase structure rules did not offer the possibility to capture obvious linguistic generalizations [Chomsky 1972]. GPSG showed that such generalizations could be expressed by employing feature bundles as representations of nodes in the phrase structure rules, as well as a meta-grammar which would generate the full set of rules from a small set of initial rules. As long as the features were non-recursive and as long as these meta-rules were not applied recursively, context-freeness of the grammar was guaranteed.

[4] The label *unification-based* was applied to LFG only much later. The distinction between functional descriptions and models is closely related to the distinction between feature descriptions and feature structures, as it is described in the HPSG literature.

[5] Cf. also [Bresnan 1978].

As it seems, in the end, GPSG was never fully implemented, nor did it have a lasting impact on the further developments. The reasons for the former was presumably that a precompilation of all rules by the meta-grammar would result in hitherto unknown degrees of space complexity [Barton et al. 1987], while the run-time application of meta-rules was never fully achieved.[6] At the same time it was conjectured that natural languages would go beyond the complexity of context-free languages ([Shieber 1985], [Culy 1985]), such that one of the main motivations and restrictions of this framework could not be further maintained.

PATR-II PATR-II [Shieber et al. 1983] in some sense marked a milestone in the development of unification-based formalisms, since it was developed as a grammar formalism which was explicitly not dedicated to some linguistic theory. Although it was lagging behind FUG in various aspects (it had a context-free backbone and its feature logic was less expressive), its simplicity and expressivity nevertheless provided an appealing platform for experimentation and the formulation of different linguistic approaches. Thus, for a long time PATR-II has been viewed as *the* canonical unification formalism. It can also be seen as one of the first "designer formalisms", i.e., as a formalism which computational linguists could use without having to buy a wholesale linguistic theory, and in which they could implement their own favourite linguistic assumptions or theory.

Categorial Unification Grammars – CUG/UCG The radical lexicalist approach of categorial grammars gave the further development a new twist when they were reformulated in unification-based formalisms (cf., [Uszkoreit 1986a], [Karttunen 1989], [Calder et al. 1988]). On the one hand, categorial grammars require only a minimal set of combinatory rules, which could be also expressed as phrase structure rules, on the other, they encode more information into the lexicon than any other theory. As a consequence the context-free backbone of PATR-II and other theories became to play a comparatively minor role, while efficient and compact techniques for the representation of the lexicon became of overall importance. One of the most important features of categorial grammars and their extensions, e.g., Combinatory Grammar [Steedman 1987], is probably that the powerful lexical operations and the absence of any rigid constituent structures allow to capture phenomena, such as non-constituent coordination, which so far defy most other theories and frameworks.

Stuttgart Typed Unification Formalism – STUF While PATR-II had only some abbreviatory mechanisms in the form of templates, it soon became clear that more powerful devices are required for a controlled grammar development

[6] Cf. however [Shieber 1986].

and grammar engineering, as for example sophisticated lexical rules as well as extensions of the feature logics by types or sorts. This demand was partly satisfied by some extensions made in UCG, and it also led to the development of one of the first grammar formalisms combining types with feature unification, namely STUF [Bouma et al. 1988].

Even if the first grammars developed in STUF were in the categorial grammar style, it was again not designed as a theory-specific formalism but rather as a general purpose linguistic description language, which in later years was also employed to implement HPSG-style grammars.

Head Driven Phrase Structure Grammar – HPSG The framework of HPSG in its original and also its current version, attempted to combine many of the features of the formalisms and frameworks we described above.

Like most previous theories HPSG stressed the independence of the description language and the linguistic theory, choosing a typed feature logic with functional constraints as the basic formalism. Similar to unification style categorial grammars it dispensed with a context-free backbone altogether and showed how even complex phrase structures could be described by the same formal device of feature structures. Same as LFG it stressed the distinction between feature descriptions and models satisfying these descriptions, developing this distinction further into a linguistic model of (structuralist) *signs* and their description.

On the linguistic side, HPSG adopted from categorial grammar the mechanisms for complementation, from LFG a large part of the control theory, from GPSG the basic treatment of unbounded dependencies, and from the Government and Binding Theory some assumptions about anaphoric binding, often challenging these approaches and taking them some steps further. However, the major achievement of HPSG appears to be not only that it managed to combine valuable insights from previous frameworks, but that a coherent and homogeneous theory evolved as a result which, with all its flaws and deficiencies, provided a stable basis for further development.

Typed Feature Structures – TFS Although it quite obviously built on previous work, the TFS system [Emele and Zajac 1990] was the first implemented formalism in the computational linguistics community, which attempted to provide a basis for implementing HPSG-style grammars more or less directly following the specifications of [Pollard and Sag 1987]. Thus, it offered not only a typed feature formalism in the spirit of HPSG but also relied on type deduction as the only means of derivation, which appeared to be the only derivation algorithm intended by the authors. The disadvantage of this implementation was that it lacked any further control devices and thus turned out to be quite inefficient for any grammars of a reasonable size.

Comprehensive Unification Formalism – CUF While TFS concentrated on types and type deduction, CUF implemented the second concept that HPSG envisaged in the basic formalism, namely the usage of (more or less) arbitrary constraints over variables [Dörre and Dorna 1993]. Although CUF does not include types as a primary descriptive device, it presumably belongs to the most expressive linguistic formalisms to date. Nevertheless it appears to be one of its weak points that the control issue is not yet solved in a satisfactory manner.

Type Description Language – TDL The $\mathcal{TDL}$ formalism[7] has been developed and gradually improved over the recent years to represent one of the most powerful languages for typed feature logics in the linguistic community. It incorporates multiple inheritance, open and closed domains, hierarchies over (atomic) sorts and typed feature structures, disjunction, negation, different kinds of incompatibility declarations, and many other concepts, which we will describe in more detail below in section 2.3.

1.2.3 Conclusion

In the previous sections we tried to sketch some aspects and tendencies in the development of constraint-based frameworks which appear relevant to our work described below. Thus, we consider it by now to be state of the art that a linguistic theory be clearly separated from the specification language or formalism. On the formal side we see a development from more constrained and linguistically motivated formalisms to more expressive and unrestricted description languages. These languages typically integrate type or sort logics with feature logics, and often allow functional or relational constraints as an extension. Context-free backbones as a descriptive device appear to have become less crucial, although this does not necessarily preclude that they are compiled from the formalism in some way or other as a means for imposing further control on derivations.

In the next sections we will attempt to put HPSG into a perspective in terms of different levels of abstraction, trying to show how some of the different frameworks we sketched can be related as implementations of these different levels.

[7] Cf. [Krieger and Schäfer 1994b], [Krieger and Schäfer 1994c], [Krieger and Schäfer 1994a], [Krieger 1995], [Krieger and Schäfer 1995].

1.3 Grammar Models and Layers of Abstraction

In this section we will discuss the relation between formalisms and linguistic theories from the perspective of levels of abstraction. We outline how this structure can be applied to HPSG and how it relates to the discussions of the formalism, the theoretical and empirical issues of HPSG and our fragment of German grammar below.

Computational models of language or of a specific language have often suffered from a mixture (or even confusion) of implementational aspects and theoretical or empirical linguistic issues. That is to say in the least that it is often not clear where the boundary between essential linguistic assumptions and matters of formalization and implementations is to be drawn. Of course, it must be granted that even for those frameworks which attempt to make these relations explicit, it is sometimes not quite clear up to what degree an implemented grammar is an instance of a particular framework. Confusions of this kind often result in statements such as “The LFG formalism does not allow co-indexation between feature paths of length greater than two.”, “This grammar implementation is not an HPSG, because it does not separate immediate dominance and linear precedence rules.”, etc.

Even if such statements are occasionally true, it is always worthwhile to try to identify the explicit assumptions made within a framework and their relations to different levels of abstraction. One such model of layers of abstractions is given in fig. 1.2. Ideally the relation between these different layers is one of inclusion on the level of expressivity and one of abstraction on the level of linguistic detail and empirical assumptions.

1.3.1 Computational and Grammar Formalism

At the bottom of this picture we find a distinction between *computational formalisms* and *grammar formalisms*. Of course, the notion of a computational formalism makes sense only if a formalism has been developed as a multi-purpose formalism, which is not committed to a specific grammar framework or which is not even necessarily constrained to linguistic applications. The formalisms listed on the left-hand side of fig. 1.1 typically fulfill these requirements. So, for example, one of the first LFG implementations other than the original, built on a DCG Prolog implementation.[8] As mentioned, PATR-II and STUF were used to implement different types of categorial grammars. TFS, CUF, TDL and other languages served as the basis of implementations of HPSG.

[8] Cf. [Eisele and Dörre 1986].

Individual Grammar	Head-Final-Rule, Lexical Entries, etc.
Grammar Theory	Head-Feature-Principle, Semantics Principle, Subcategorization Principle
Grammar Model	Principles $\wedge$ Universal Rules $\vee$ Lang. Spec. Rules $\vee$ Language Specific Lexical Items
Grammar Formalism	Typed Feature Logics with Relational Constraints
Computational Formalism	e.g., TDL, TFS

Figure 1.2: Layers of Abstraction

Typically, a computational formalism is a 'lower' language than the grammar formalism, such that the implementation of a grammar formalism in a computational formalism may require some compilation. One example of such a case we find in the DCG implementation of LFG, where the syntax and semantics of the grammar formalism was very close to the original LFG, while the compilation into DCG resulted in an elimination of certain descriptive devices, such as Kleene-star, left recursions, etc. Concepts which are not a priori given by the computational formalism may have to be programmed in addition or require some special compilation. Again, an example would be the non-monotonic parts of LFG, such as existential constraints or negation, which require a special treatment in a DCG.

This raises the question whether a computational formalism must always be more powerful than or at least as powerful as the grammar formalism implemented in that formalism. For example, there have been long discussions whether LFG can be expressed in PATR-II, which was always denied by Kaplan, mainly on the basis that PATR-II does not comprise non-monotonic devices. To date this question is still unresolved, since so far nobody ever showed whether or not these non-monotonic devices can be eliminated or compiled out into straightforward unification.

One aspect of this issue is also whether one should not voluntarily accept the restrictions of a computational formalism for the sake of a more efficient implementation. However, such a strategy can be advocated only as long as the concepts higher up in the hierarchy, i.e., in the grammar model and theory, can be expressed in a comparable way by using less powerful devices. Such a situation we typically find in cases, where the grammar formalism is very close to the computational formalism, but where the computational formalism does not fully implement the whole range of possibilities or concepts. Examples here are most of the above-mentioned HPSG implementations, which either cannot foresee all possible relational constraints or which do not provide the full range of typed feature logics. We will come back to this very central issue in our discussions of the TDL formalism and the implementation of principles below.

Leaving aside these issues, the *grammar formalism* nevertheless has to be seen as the major tool in which the grammar writer expresses his linguistic descriptions and generalizations. While a computational formalism may or should be devised primarily along the requirements of efficiency, this level of abstraction should be tailored towards the need of grammar writing, providing as much expressivity and support to the specific linguistic requirements as possible.

Thus, for HPSG we would identify the grammar formalism with a standard typed feature logics with relational constraints (which may but need not involve type deduction as the basis for derivations). In the case of LFG we find context-free phrase structure rules with feature annotations. For categorial grammar, the grammar formalism is defined by the combinatorial principles of functional application, augmented by functional composition and type raising in some cases. All of these grammar formalisms have been given a declarative syntax and semantics which can serve as the common basis both for the linguistic descriptions and the implementation of the derivation algorithm.

1.3.2 Grammar Model

By the notion of a grammar model we understand the general architecture of a linguistic framework, or more specifically the constraints that the set of grammars comprised by this framework should satisfy. These are the invariants of a linguistic framework that make up its essence and that make predictions on how this framework attempts to model the empirical observations on language.

By definition, the grammar model has to be a very abstract concept which cannot make predictions beyond a certain level of detail. For example, for HPSG the grammar model involves the basic assumption that the description of all linguistic objects should be based on the notion of a *sign*, which relates the form to the meaning. The descriptions of signs have to satisfy a set of constraints which is composed from language universal and language specific constraints. Specifically, it is assumed that there are universal principles, which always have to be satisfied, and that there are some universal rule schemata from which the language specific rule schemata or instances of phrasal configurations are drawn and further specified. The description of lexical signs is assigned to the language specific part, although one could envisage an analogous distinction of universal and language particular lexical rules.

Essential to the grammar model of LFG is a basic distinction between categories and grammatical functions, which are assigned to a c-structure and an f-structure level of abstraction. While the categories and in particular the different forms of phrase structure configurations are assumed to be language specific, the grammatical functions are assumed to be taken from a universal set of functions which applies across different languages. Later extensions of the model included the assumption of additional level of projections for the semantic or σ-structure.

As to the relation between the grammar model and the grammar formalism, again HPSG and LFG provide interesting examples. The LFG model postulates the existence of two different syntactic and (at least) one semantic level of abstraction (c-, f- and σ-structure), of which only the latter two are described within the same formal device of feature constraints. In contrast the different levels of the entire model of HPSG are assumed to be described by the same formal device of typed feature structures. However, it is not very difficult to see how LFG could be expressed in a pure typed feature formalism or how a context-free backbone could be compiled out of HPSG.

1.3.3 Grammar Theory

While the grammar model defines what kind of "boxes" there are in a grammar, it is the task of the grammar theory to specify what kind of information should go into those boxes.

To use the example of HPSG again, the model of HPSG postulates the existence of universal principles and defines the possible relation and interaction between these principles and language particular constraints. A particular theory of HPSG has to define the content of these principles, the set of rule schemata, the ontological concepts, such as the notion of head, the different kinds of possible grammatical relations, relations such as agreement or government etc.

Similarly, a theory of LFG has to give specific definitions of the set of grammatical functions, of the set of phrase structure categories and what the possible assignments of functions to categories are.

It is quite obvious from the history that there never existed *the* theory of HPSG, of LFG or of any other framework. Theories always change in the course of time (as long as they are being alive) and very frequently different theories compete with each other. If we compare the different versions of HPSG, which have been proposed in [Pollard and Sag 1987] and [Pollard and Sag 1994], many of the definitions of individual rule schemata and principles have changed quite radically, while the overall assumptions of the grammar model have only undergone some further refinement but are in their essence still maintained.

1.3.4 Individual Grammar

Individual grammars for specific languages can be considered as instances of a grammar theory. Ideally, these grammars should follow the predictions made by the theory, just as it was suggested above in the model of HPSG: language particular generalizations should be more specific instantiations of universal generalizations set up in the theory.

Of course the predictive power of a theory cannot be measured exclusively in terms of what predictions can be maintained when tested against a new language, since the less predictions a theory makes, the less likely these predictions are going to fail. It will also depend on the general amount of 'guidance' a theory provides for the specification of a new grammar.

On the other hand, a theory cannot be too restrictive as long as it has not been tested against a sufficiently large and representative number of languages. In particular, the setup of the theory should be such that it can be modified locally when necessary without the entire system breaking down. A theory which is too restrictive and in which the interdependencies of different modules of the theory are too tight could easily enforce or at least provoke inadequate analyses for new phenomena or languages.

Thus, the value of a theory shows itself in the degree of support it provides for the development of a new language specific grammar, both in terms of the predictions it makes and in terms of the variability it permits.

While there may be an ideal state of a theory, it is quite clear that in practice most theories grew out of the description of one particular language (for example English), which has consequently been generalized and applied to other languages. Therefore, most theories suffer in the beginning from the fact that they have been tailored towards a single language or at least a single type of language. Only by incorporating insights from applications to other languages can a theory be made truly general.

1.3.5 Results Related to the Layered Model

In the previous sections we sketched the structure of linguistic frameworks in terms of levels of abstraction. On the basis of this layered model, we describe in this section at what levels this thesis attempts to make contributions.

Grammar Formalism

In the domain of the grammar formalism we will build on an existing typed feature logic, namely the formalisms of $\mathcal{TDL}$ (Type Description Language). We take $\mathcal{TDL}$ as the basis of our descriptions not only because the grammar fragment described has been implemented in $\mathcal{TDL}$, but also because $\mathcal{TDL}$ to us appears to be one of the most powerful and expressive typed feature formalisms in the area of computational linguistics.

$\mathcal{TDL}$ differs from other formalisms, such as ALE or TFS, among many other things in that it allows open domains in the type hierarchy, multiple inheritance at all levels of embedding, as well as full negation and disjunction of types and features. In chapter 2, where we describe the formalism, and we also discuss a potential extension $\mathcal{TDL}$ by *cross-classification* which could allow a more straightforward implementation of the notion of principles in HPSG.

Grammar Model

Although we accept most of the basic assumptions of the grammar model of HPSG, we challenge some of them where they appear too constrained or void. Thus, in chapterHPSG, we discuss the notion of universal and language-particular grammar and outline an alternative, more flexible view building on the notion of generalizations as they are expressible in typed feature logics.

Grammar Theory

At the level of the grammar theory of HPSG, we propose a new approach concerning the status and form of phrase structure schemata and grammatical relations on the basis of a discussion of the notion of *head*. We also present a novel and alternative treatment of functional categories, which allows us to simplify the existing set of rule schemata and principles.

Since the notion of *head* is one of the central notions in HPSG, we discuss this concept from different points of view in order to identify some defining properties of heads. These aspects concern the assumption and motivation of head *features* and the associated domains of locality, but also the notion of *head* vis á vis to notions such as *functor* or *selector*, which play a role in the determination and definition of grammatical relations.

Building on the notion of grammatical relations evolving out of the discussion of *heads*, we outline an alternative conception of phrasal configurations. The revised standard theory of HPSG[9] is based on the assumption that the constituent structure of natural languages can be described by up to six (or possibly seven) different rule schemata, which define the grammatical functions, selectional relations, the phrasal or lexical status of constituents, and other aspects of phrase structure. We show that without any apparent loss of generalization out of these seven rule schemata five can be reduced to a single basic grammatical head-complement-relation, which comes in the form of a head-subject- and a head-complement-relation. Next to these two only head-adjunct- and head-filler-relations are acknowledged as the basic building blocks of the grammar theory. This goes together with the assumption that only three basic types of selectional relations are necessary.

We achieve this reduction in three steps. We first reduce the three rule schemata in HPSG which define different types of complementation, namely head-subject-, head-complement- and head-subject-complement-structures. By generalizing the first and redefining the second schema to a binary structure, in which the head is combined with only one complement at a time, we open the possibility to treat the third schema as a derivative of the former two. Far from a simple formal reshuffling of information, this changed setup allows to define a basic universal grammar from which language particular grammars can be derived by means of additional specification and parametrization. We demonstrate how these redefined two basic schemata can be employed to capture the distinction between configurational languages and languages with a more relaxed word order, how unary rule derivations can be eliminated, and that linear precedence constraints building on the distinction between lexical and phrasal categories have an equally simple equivalent in our approach.

Second, we show how the head-marker-schema (as well as the head-specifier-schema postulated in some unpublished versions of HPSG) can be eliminated altogether, if determiners and complementizers are treated as functional heads, which take the nominal phrase or clausal body respectively as complements. This analysis is based on the assumption that head features are split into two sets, major and minor features, where the former determine information such as part

[9] This refers to the revisions of [Pollard and Sag 1994] in chapter 9.

of speech, agreement, etc., whereas the latter serves to define a property which we call functional completeness. Functional heads are then treated as specific heads which provide the head features to the phrasal projection, but adopt the major part of the head features from their complements. Since all of the relevant information can be specified in the lexicon, it is almost needless to say that structural principles such as the Spec-Principle and the Marker-Principle do not have a place in this approach either.

Third, as it was suggested to employ the original HPSG head-subject-complement-schema for the analysis of free word order languages such as German on the basis of a flat structure, we show how the relevant phenomena can be captured more adequately by means of a binary branching structure requiring only our basic head-complement relations. For this purpose we introduce an empty verbal functional head for the analysis of V-initial clauses, which takes the finite verb as a lexical complement. Through the constraints determining functional completeness the complementary distribution of complementizers and the initial position of the finite verb falls out as a natural consequence.

Individual Grammar

The suggested modifications of the theory of HPSG are partly also motivated by the language-specific analysis proposed for a fragment of German. This concerns above all the last two issues mentioned, the treatment of functional heads and the treatment of free word order languages. Thus, the core of this fragment is formed by phenomena involving functional heads in the domain of nominal phrases as well as verbal projections.

A more general question underlying our treatment of determiners and determinerless constructions in German concerns the relation between empty categories and unary rules on the one hand, and disjunction and underspecification on the other. In our analysis of determinerless constructions (mass nouns and bare plurals) in German we explore these different approaches and propose a solution which discards with the first three, complex representations and show how the same results can be achieved by pure underspecification. Thus, our analysis of constructions such as *Bücher*, *Weine*, vs. **(das) Buch*, *(der) Wein*, requires neither the application of a unary rule to derive a maximal phrasal projection, nor the introduction of an empty determiner, nor the postulation of disjunctive subcategorization, but operates with simple underspecification in the lexicon.

An issue related to this problem is the question of intermediate projections and adjunction of attributive adjectives. We show that the traditional notion of maximal projections has to be considered as defective and that an adequate treatment of attributive adjunction (in German as well as in other languages) will enforce the introduction of unary rules or (formally equivalent) empty categories in the traditional HPSG approaches. In contrast, our analysis building on the split of head features provides a natural solution which allows to capture the phenomena on the lexical level without any such devices. For those who do not wish to follow our approach, we nevertheless provide a solution which allows to capture the phenomenon also within the boundaries of the traditional HPSG approach by generalizing the head-marker-relation and modifying the Spec-Principle.

While the above two aspects could have some repercussions which are relevant for the grammar theory of HPSG, there are also a series of issues which are more closely related to German as a specific language. The first phenomenon involves empty or missing noun heads in nominal phrases, which could be treated either as cases of ellipsis, where an empty head noun is postulated, or as a case of pronominalization. We first discuss in detail the descriptive data providing the motivation for postulating empty nominal heads functioning as anaphoric heads similar to English *one*-Anaphora. Although it is acknowledged that there may be also some alternative motivation for treating certain determiners in isolation as pronominalized categories equivalent to maximal nominal phrases in the lexicon, we opt for a solution which generally introduces an empty nominal head for these constructions. We also specify the constraints which are necessary to exclude the ungrammatical combinations of the apparently optional occurences of determiners, adjectives and nouns.

Throughout the treatment of nominal phrases the agreement phenomena as well as declension of determiners, adjectives and nouns, which are specific to German, play an important role. We show that these two phenomena can be appropriately accounted for by making use of the distinction between major and minor features again. While agreement is captured on the level of major features which are typically shared throughout nominal phrases, declension is viewed as a kind of government relation which is described on the level of minor features. Specifically, we claim that the formation of declension classes on the basis of the type or stem of the determiner fails to capture important generalizations and that these generalizations can be described more adequately on the basis of the opposition between inflected and non-inflected instances of determiner forms. We show how the frequently postulated three declension classes (strong, weak and mixed) can be reduced to a binary distinction and that the description of declension phenomena requires only a small set of simple axioms, some of which cut across the inflectional paradigms of determiners in a very natural way. The basic assumption are that all maximal projections of nominal phrases are marked for strong declension, that all strongly inflected determiner govern their complements to be weakly inflected, and that uninflected determiner forms co-index their declension values with their complements. Numerals, which have a status somewhere in between determiners and adjectives are induced as additional evidence for our analysis of agreement and declension. While in principle independent of the issues of head-complement-relations, the declension phenomena also serve to corroborate the analysis of determiners as functional heads which govern the declension values of their nominal complements.

Our theory of phrase structure configurations and functional heads is finally applied to phenomena involving verbal projections and clause structure. We first investigate word order phenomena in the verbal domain which provide arguments against a flat clausal structure for complements. The alternative binary rules proposed for complementation allow to capture the free sequential order of adjuncts and complements in a more adequate way. We then show that the assumption that complementizers are functional heads and that finite verbs are inherently marked as functionally incomplete will lead to a straightforward analysis of V-final clauses. V-initial clauses are then discussed as a problem which illustrate the close dependency and effects arising from the interaction among the position of the head, the subcategorization mechanism employed and the type of the resulting phrase structure. Specifically, we argue that the subcategorization mechanisms of HPSG, a similar phrase structure for V-final and V-initial structures and the assumption that finite verbs are analyzed as heads *in situ* cannot be maintained simultaneously. We therefore suggest an alternative analysis for V-initial constructions on the basis of a mechanism involving an empty functional head. One of the important advantages of this approach over movement-based analyses and others is that it also allows to capture the complementary distribution of complementizers and the finite verb in an initial position.

Chapter 2

Grammar Formalism

2.1 Introduction

In this chapter we describe the grammar formalism that we employ as a description language for the grammar fragment. This formalism is closely modelled on the Type Description Language (TDL) developed by Krieger and Schäfer,[1] making use of the unification engine UDINE. For readers who are familiar with standard typed feature logics, this chapter will contain little new information, except for the notation we employ and some extensions that TDL provides. It may also help to understand the potentially different kind of perspective on typed feature structures that the open world assumption of TDL allows.

To illustrate the role of the formalism in a linguistic framework we use the well-known diagram in fig. 2.1 on the relationship between phenomena, models of phenomena and formal representations [Pollard and Sag 1994][7f].

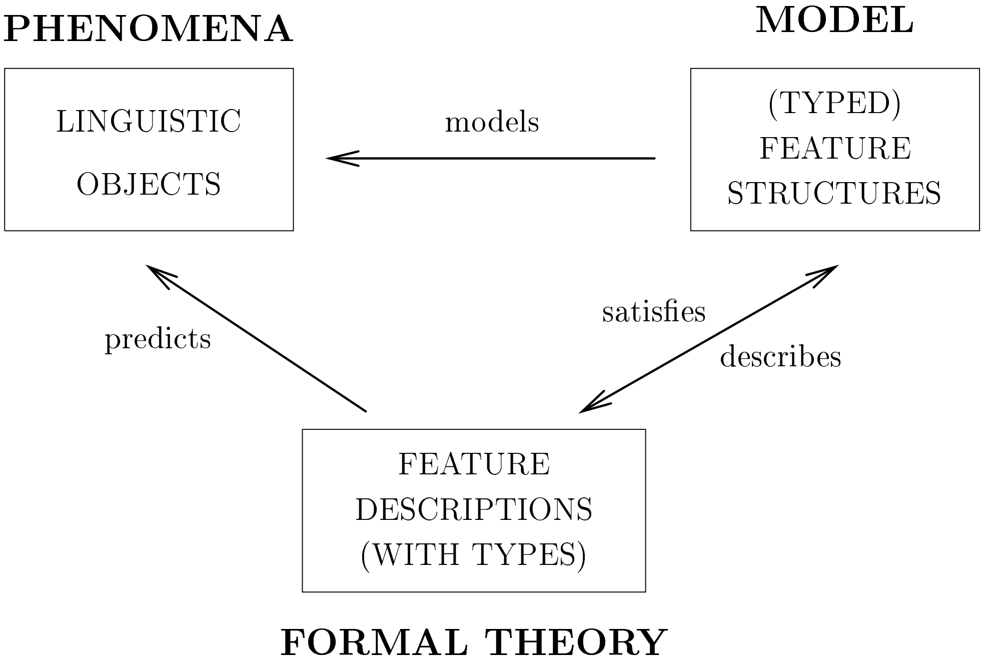


Figure 2.1: Formal Theories, Models and Phenomena

Under this view, formal theories describe models of empirically observable phenomena. Applied to linguistic frameworks such as HPSG, the formal theory is

[1] Cf. [Schäfer and Krieger 1992], [Backofen and Krieger 1993], [Krieger and Schäfer 1994b], [Krieger and Schäfer 1994c], [Krieger and Schäfer 1994a], [Krieger and Schäfer 1995], [Krieger 1995].

instantiated by a feature logic augmented with types or sorts, which describe typed feature structures. Typed feature structures on the other hand are models which satisfy descriptions. Thus, a grammar formalism will cover aspects of the formal theory as well as the model.

2.2 Feature Logic

In describing linguistic objects by means of a feature logic we assume the standard distinction of *feature descriptions* and *feature structures* (cf. fig. 2.1). Feature descriptions consist of sets of constraints in the form of attribute value pairs. Feature descriptions describe feature structures, which can be interpreted, for example, as labelled graphs. Feature structures are models of linguistic objects and satisfy feature descriptions.

2.2.1 Feature Descriptions

Feature descriptions consist of sets of constraints connected by the boolean operators "&" (conjunction), "|" (disjunction), and "¬" (negation). Constraints are represented as attribute value pairs, where the values can be atoms, variables, attribute value pairs or functional constraints over variables.

Since we are mainly concerned with syntactic aspects in our grammar fragment, we do not foresee sets as possible values. Where set-like concepts are required, we expect them to be expressed by list operations.

To exclude the simultaneous usage of an atom as a value and an attribute, as it is possible in LFG, we exclude the possibility that attributes are represented by variables.

We will use graphical representations for feature descriptions whenever possible. Conjunction, disjunction, and negation are represented by square "[...]" brackets, curly "{...}" brackets and the "¬" symbol respectively.

$$(1)\quad \begin{bmatrix} \texttt{CONJ1}\ \{\textit{disj1, disj2}\} \\ \texttt{CONJ2}\ \neg \textit{neg} \end{bmatrix}$$

For attributes we conventionally use capital letters (ATTR), for atomic values small italics (*atom*). Variables or indices representing token identical or coreferent values are given as boxed integers $\boxed{1}$.

The "$\top$" symbol represents the empty constraint set (denoting the universe of objects) and is notated as "[ ]" in graphical representations. The "$\perp$" symbol stands for an inconsistent constraint set (denoting an empty set of objects).

Lists as values of attributes are marked by angled brackets with some abbreviatory conventions for (a) empty lists $(<\ >)$, (b) lists with one element $(<[\ ]>)$, (c) lists of variable length, $n \geq 0$ $(< ... >)$ and (d) open lists of minimal length n (e.g., $n \geq 2$, $<[\ ],[\ ],...>$). These notations are considered to be equivalent to the values in the following representations respectively:

(2) (a) $\left[\texttt{LIST}\ *end\right]$

(b) $\left[\texttt{LIST}\left[\begin{array}{ll}\texttt{FIRST} & [\ ] \\ \texttt{REST} & *end*\end{array}\right]\right]$

(c) $\left[\texttt{LIST}\left\{\begin{array}{l}\left[\begin{array}{ll}\texttt{FIRST} & [\ ] \\ \texttt{REST} & [\ ]\end{array}\right] \\ *end*\end{array}\right\}\right]$

(d) $\left[\texttt{LIST}\left[\begin{array}{ll}\texttt{FIRST} & [\ ] \\ \texttt{REST} & \left[\begin{array}{ll}\texttt{FIRST} & [\ ] \\ \texttt{REST} & [\ ]\end{array}\right]\end{array}\right]\right]$

Functional constraints over values are conditions over variables occurring in the feature descriptions. Functional constraints can be specified as values of attributes within a feature description (3) or as separate conditions making reference to the respective variables (4).

(3) $\left[\begin{array}{ll}\texttt{FIRST-ARG} & \boxed{1} \\ \texttt{SECOND-ARG} & \boxed{2} \\ \texttt{RESULT} & \text{func-constr } (\boxed{1},\ \boxed{2})\end{array}\right]$

(4) $\left[\begin{array}{ll}\texttt{FIRST-ARG} & \boxed{1} \\ \texttt{SECOND-ARG} & \boxed{2} \\ \texttt{RESULT} & \boxed{3}\end{array}\right]$

COND: $\boxed{3} = \text{func-constr } (\boxed{1}\ ,\ \boxed{2}\ )$

We will prefer the latter representation which supports the view that unification operations over feature descriptions can be applied independent of the evaluation of functional constraints.

As a specific extension of standard unification formalisms we employ the concept of distributed or named disjunctions, which allow the specification of embedded disjunctions without the computation of disjunctive normal forms (DNF). On the basis of this device two or more disjunctions with the same number of disjuncts and appearing as values of different paths can be given a unique identifying name (marked as \$n), which will cause the n-th disjunct in each disjunction to fail or succeed simultaneously.[2]

$$(5)\quad \begin{bmatrix} \texttt{PATH1}\ \$1\{\textit{disj1, disj2}\} \\ \texttt{PATH2}\ \$1\{\textit{disj3, disj4}\} \end{bmatrix} \sqcap \begin{bmatrix} \texttt{PATH1}\ \textit{disj2} \end{bmatrix} \equiv \begin{bmatrix} \texttt{PATH1}\ \textit{disj2} \\ \texttt{PATH2}\ \textit{disj3} \end{bmatrix}$$

2.2.2 Feature Structures

We view feature structures as models of sets of entities (e.g., abstract primitive properties, sets of properties, linguistic expressions etc.) which satisfy constraint sets or feature descriptions.

A feature structure satisfies a constraint iff the set of entities modelled by that feature structure has the property described by the constraint. The empty constraint set $\top$ or [ ] is satisfied by all feature structures and thus denotes all entities of the universe. The conjunction of two constraints which assign two distinct atomic values to the same attribute is satisfied by no feature structure,[3] and corresponds to the empty set of entities ($\perp$).

A feature structure Fc satisfies a conjunctive constraint ϕ & ψ iff it denotes the intersection of the sets of entities denoted by F_ϕ and F_ψ. A feature structure Fd satisfies a disjunctive constraint $\phi \mid \psi$ iff it denotes a set of elements in the union of the sets of entities denoted by F_ϕ and F_ψ. A negative constraint $\neg\phi$ is satisfied by all feature structures which denote elements that are not in the set denoted by the feature structure F_ϕ.

Two sets of constraints ϕ and ψ can be conjoined by means of a unification operation $\sqcap$ which yields as a result the minimal constraint set which denotes the intersection of the constraint sets ϕ and ψ.

A Note on Disjunctions

The described formalism will allow almost free combinations of the boolean operators and variables. However, there is a restriction on the usage of variables in combination with disjunction to the effect that identical variables have to occur in the same disjunctive context. We briefly elaborate on this restriction since it also motivates the introduction of the concept of distributed disjunctions into the formalism.[4]

[2]For a more elaborate description and motivation see below in this section.

[3]See however below our usage of sorts as a means to represent hierarchies of atoms in a type system.

[4]Cf. also [Karttunen 1984] for a description of the problem.

To begin with a simple example, if we want to say that in a feature structure an attribute `A` should have the same value as either a path `B C` or a path `B D` we cannot specify this as (6):

$$(6)\quad \begin{bmatrix} \texttt{A} & \boxed{1} \\ \texttt{B} & \left\{ \begin{bmatrix}\texttt{C} & \boxed{1}\end{bmatrix}, \begin{bmatrix}\texttt{D} & \boxed{1}\end{bmatrix} \right\} \end{bmatrix}$$

This specification could lead to inconsistencies and incomplete results depending on the order in which further unifications are carried out. If we unified this structure stepwise with two structures assigning different values to `C` and `D` as in (7) and (8), this will lead to the result in (8).

$$(7)\quad \begin{bmatrix} \texttt{A} & \boxed{1} \\ \texttt{B} & \left\{ \begin{bmatrix}\texttt{C} & \boxed{1}\end{bmatrix}, \begin{bmatrix}\texttt{D} & \boxed{1}\end{bmatrix} \right\} \end{bmatrix} \sqcap \begin{bmatrix}\texttt{B} & \begin{bmatrix}\texttt{C} & +\end{bmatrix}\end{bmatrix} \equiv \begin{bmatrix} \texttt{A} & + \\ \texttt{B} & \left\{ \begin{bmatrix}\texttt{C} & +\end{bmatrix}, \begin{bmatrix}\texttt{C} & + \\ \texttt{D} & +\end{bmatrix} \right\} \end{bmatrix}$$

$$(8)\quad \begin{bmatrix} \texttt{A} & + \\ \texttt{B} & \left\{ \begin{bmatrix}\texttt{C} & +\end{bmatrix}, \begin{bmatrix}\texttt{C} & + \\ \texttt{D} & +\end{bmatrix} \right\} \end{bmatrix} \sqcap \begin{bmatrix}\texttt{B} & \begin{bmatrix}\texttt{D} & -\end{bmatrix}\end{bmatrix} \equiv \begin{bmatrix} \texttt{A} & + \\ \texttt{B} & \begin{bmatrix}\texttt{C} & + \\ \texttt{D} & -\end{bmatrix} \end{bmatrix}$$

Changing the order of the unifications, i.e., unifying (6) first with $[\,\texttt{B}\ [\ \texttt{D}\ -]\,]$ and then with $[\,\texttt{B}\ [\ \texttt{C}\ +]\,]$ will lead to a different result. Even if we first unified the respective second structures in (7) and (8) with each other and only then with (6), the result would still depend on which of the disjuncts in (6) is considered first. At the very best we will obtain two separate results corresponding to a DNF.

$$(9)\quad \begin{bmatrix} \texttt{A} & \boxed{1} \\ \texttt{B} & \left\{ \begin{bmatrix}\texttt{C} & \boxed{1}\end{bmatrix}, \begin{bmatrix}\texttt{D} & \boxed{1}\end{bmatrix} \right\} \end{bmatrix} \sqcap \begin{bmatrix}\texttt{B} & \begin{bmatrix}\texttt{C} & + \\ \texttt{D} & -\end{bmatrix}\end{bmatrix} \equiv \begin{matrix} \begin{bmatrix} \texttt{A} & + \\ \texttt{B} & \begin{bmatrix}\texttt{C} & + \\ \texttt{D} & -\end{bmatrix} \end{bmatrix} \\ \begin{bmatrix} \texttt{A} & - \\ \texttt{B} & \begin{bmatrix}\texttt{C} & + \\ \texttt{D} & -\end{bmatrix} \end{bmatrix} \end{matrix}$$

A correct and complete result, which is independent of the order of unifications, we will obtain in this simple case, if there are no coreferences between the disjuncts and if the value of `A` is specified as a disjunction over two independent variables. In (10) we repeat only the operation in (7) to illustrate the difference.

$$(10)\quad \begin{bmatrix} \texttt{A} & \left\{ \boxed{1}\,,\ \boxed{2} \right\} \\ \texttt{B} & \left\{ \begin{bmatrix}\texttt{C} & \boxed{1}\end{bmatrix}, \begin{bmatrix}\texttt{D} & \boxed{2}\end{bmatrix} \right\} \end{bmatrix} \sqcap \begin{bmatrix}\texttt{B} & \begin{bmatrix}\texttt{C} & +\end{bmatrix}\end{bmatrix} \equiv \begin{bmatrix} \texttt{A} & \left\{ +,\ \boxed{2} \right\} \\ \texttt{B} & \left\{ \begin{bmatrix}\texttt{C} & +\end{bmatrix}, \begin{bmatrix}\texttt{C} & + \\ \texttt{D} & \boxed{2}\end{bmatrix} \right\} \end{bmatrix}$$

However, the simple restriction that disjuncts must not contain coreferent variables will not lead to correct results in all configurations. It can be shown by a more complex example that only the stronger version of *distributed* disjunctions can guarantee correctness (unless disjunctive constraints are compiled out into DNFs).

Suppose that we want to express that either the value of a path `A C` is coreferent with a path `B E`, or that a path `A D` must have the same value as a path `B F`. At first sight, one might expect that the description in (11) will capture this constraint.

$$(11)\quad \begin{bmatrix} \mathtt{A} & \left\{ \begin{bmatrix}\mathtt{C}\ \boxed{1}\end{bmatrix}, \begin{bmatrix}\mathtt{D}\ \boxed{2}\end{bmatrix} \right\} \\ \mathtt{B} & \left\{ \begin{bmatrix}\mathtt{E}\ \boxed{1}\end{bmatrix}, \begin{bmatrix}\mathtt{F}\ \boxed{2}\end{bmatrix} \right\} \end{bmatrix}$$

However, this definition will yield the intended result only as one possible solution among others, which can be easily shown by unifying the different disjuncts with a structure which assigns different values crosswise to the attributes and disjuncts.

$$(12)\quad \begin{bmatrix} \mathtt{A} & \left\{ \begin{bmatrix}\mathtt{C}\ \boxed{1}\end{bmatrix}, \begin{bmatrix}\mathtt{D}\ \boxed{2}\end{bmatrix} \right\} \\ \mathtt{B} & \left\{ \begin{bmatrix}\mathtt{E}\ \boxed{1}\end{bmatrix}, \begin{bmatrix}\mathtt{F}\ \boxed{2}\end{bmatrix} \right\} \end{bmatrix} \sqcap \begin{bmatrix} \mathtt{A} & \begin{bmatrix}\mathtt{C}\ +\end{bmatrix} \\ \mathtt{B} & \begin{bmatrix}\mathtt{F}\ -\end{bmatrix} \end{bmatrix} \equiv \begin{bmatrix} \mathtt{A} & \left\{ \begin{bmatrix}\mathtt{C}\ +\end{bmatrix}, \begin{bmatrix}\mathtt{C}\ + \\ \mathtt{D}\ -\end{bmatrix} \right\} \\ \mathtt{B} & \left\{ \begin{bmatrix}\mathtt{E}\ + \\ \mathtt{F}\ -\end{bmatrix}, \begin{bmatrix}\mathtt{F}\ -\end{bmatrix} \right\} \end{bmatrix}$$

Unifying the result of (12) with the structure in (13) will cause the second disjunct in the value of `A` and the first disjunct in the value of `B` to fail.

$$(13)\quad \begin{bmatrix} \mathtt{A} & \left\{ \begin{bmatrix}\mathtt{C}\ +\end{bmatrix}, \begin{bmatrix}\mathtt{C}\ + \\ \mathtt{D}\ -\end{bmatrix} \right\} \\ \mathtt{B} & \left\{ \begin{bmatrix}\mathtt{E}\ + \\ \mathtt{F}\ -\end{bmatrix}, \begin{bmatrix}\mathtt{F}\ -\end{bmatrix} \right\} \end{bmatrix} \sqcap \begin{bmatrix} \mathtt{A} & \begin{bmatrix}\mathtt{D}\ +\end{bmatrix} \\ \mathtt{B} & \begin{bmatrix}\mathtt{E}\ -\end{bmatrix} \end{bmatrix} \equiv \begin{bmatrix} \mathtt{A} & \left\{ \begin{bmatrix}\mathtt{C}\ + \\ \mathtt{D}\ +\end{bmatrix}, \bot \right\} \\ \mathtt{B} & \left\{ \bot, \begin{bmatrix}\mathtt{E}\ - \\ \mathtt{F}\ -\end{bmatrix} \right\} \end{bmatrix} \equiv \begin{bmatrix} \mathtt{A} & \begin{bmatrix}\mathtt{C}\ + \\ \mathtt{D}\ +\end{bmatrix} \\ \mathtt{B} & \begin{bmatrix}\mathtt{E}\ - \\ \mathtt{F}\ -\end{bmatrix} \end{bmatrix}$$

The disjuncts surviving the operation on (13) are those which have values assigned to their attributes which are exactly opposite to what one may have intended as a result. Neither does `C` have the same value as `E`, nor does `D` have the same value as `F`.

Quite obviously, the connections between the coreferent values in the original structure (10) are not strong enough, but can be broken by crosswise eliminating one of the disjuncts until there are no more coreferent values. Such an effect can be avoided, however, if we not only link values but also disjuncts to each other, as it is the case with distributed disjunctions. The same operation as in (13), repeated in (14), will lead to a failure if based on distributed disjunctions, since the n-th disjunct in every distributed disjunction with the same identifier fails if one n-th disjunct fails. In this case, in the disjunction under `A` the first disjunct, in the disjunction under `B` the second disjunct fails, which means that the entire disjunction and thus the entire structure denotes the empty set.

(14) $$\begin{bmatrix} \texttt{A}\ \$1\left\{\begin{bmatrix}\texttt{C}\ +\end{bmatrix}, \begin{bmatrix}\texttt{C}\ + \\ \texttt{D}\ -\end{bmatrix}\right\} \\ \texttt{B}\ \$1\left\{\begin{bmatrix}\texttt{E}\ + \\ \texttt{F}\ -\end{bmatrix}, \begin{bmatrix}\texttt{F}\ -\end{bmatrix}\right\} \end{bmatrix} \sqcap \begin{bmatrix} \texttt{A}\begin{bmatrix}\texttt{D}\ +\end{bmatrix} \\ \texttt{B}\begin{bmatrix}\texttt{E}\ -\end{bmatrix} \end{bmatrix} \equiv \begin{bmatrix} \texttt{A}\ \$1\left\{\begin{bmatrix}\texttt{C}\ + \\ \texttt{D}\ +\end{bmatrix}, \bot\right\} \\ \texttt{B}\ \$1\left\{\bot, \begin{bmatrix}\texttt{E}\ - \\ \texttt{F}\ -\end{bmatrix}\right\} \end{bmatrix} \equiv \bot$$

2.3 Typed Feature Structures

In this section, we describe the extension of feature logics by *types* and/or *sorts*. After a short overview, we introduce type definitions, type hierarchies and inheritance relations, as well as different types of incompatibility relations between types.

2.3.1 Overview

Whereas feature constraints describe (models of) linguistic objects in terms of their individual properties, the concept of types allows to refer to such objects as such be means of class names. Typed feature formalisms as description languages for grammars have several practical and theoretical advantages.

Type definitions and type hierarchies allow for a compact and controlled description of representations of linguistic generalizations. They allow to cluster sets of constraints into classes representing the ontological entities postulated in a theory. Through hierarchies of types over which inheritance relations are defined, important generalizations can be captured in a modular way. Subtypes inherit all feature constraints which are associated with a type and are thus subject to all generalizations defined higher up in the hierarchy.

Rather than allowing arbitrary combinations of feature constraints, type hierarchies can be used to restrict the formation of ontological classes and their intersection. An inheritance relation between a subtype and its supertype(s) can be defined only if the feature constraints associated with the subtype unify with the inherited constraints. Thus, type definitions allow to impose maximal control over specifications by excluding the inadvertent assignment of non-appropriate attribute values.

From a processing point of view, typed feature structures have mainly two advantages. Types can be declared to be incompatible by means of a closed world assumption or by other forms of partially closing type lattices. Thus, if two typed feature structures have incompatible types at their roots, an eventual failure of the unification of these two structures can be detected at the earliest time possible by simple inspection of the type lattice, without the feature unification having to be carried out.

Second, the feature constraints associated with a type can always be computed from the type definitions by expanding the types. This allows to process with unexpanded or partially expanded types, i.e., structures which are considerably smaller than the fully expanded constraint sets. The constraints associated with the unexpanded types are then filled in only after the full derivation has succeeded.

In unification algorithms where the size of the structures plays a crucial role, this can be very profitable for the treatment of those constraints which are known or expected to succeed anyway. This could be the case for example for semantic constraints, if the semantic representations do not have a filtering effect but only an interpretative function.

In the following, we will make a terminological (and formal) distinction between *types*, *sorts* and *atoms*.

In this terminology *types* are names which are associated with feature structures. This means that a set of entities which is described by a complex feature description and modelled by a feature structure is labelled by an atomic name which refers to this feature structure. Formally, the relation between a type and a feature structure is established by a type definition, which assigns a feature description to the type. Intuitively, these feature descriptions can be seen as describing the appropriate properties of elements of the class denoted by a type by means of attribute value pairs. Types are assumed to be ordered in a type hierarchy which represents the subsumption relations between the sets of entities denoted by each type.

The term *sort* we reserve to atomic feature values which are not associated with a feature structure, but which share the property of types of being organised in a hierarchical lattice representing subsumption hierarchies. Types and sorts by definition denote disjoint sets of entities.

Atoms (or constants) can be seen as a special kind of sort which could be represented in a flat lattice, such that no atom is subsumed by any other object (atom, sort or type) nor subsumes any other object.

2.3.2 Type Definitions

Type definitions define what properties (described by feature constraints) are associated with an element in the set of objects denoted by the type. These definitions can be given in the form of appropriateness definitions or more generally in terms of arbitrary feature constraints.

Appropriateness Definitions

By means of appropriateness definitions it can be defined what attributes are appropriate for a given type and what types are appropriate for the values of these attributes.[5]

[5] Cf. also [Carpenter 1992][85ff].

To define the appropriateness of a feature for a given type, a function **app** can be assumed which specifies what attributes are defined ($\downarrow$) for this type (15) and what the possible ranges of values for these attributes are, expressed again by means of a type (16).

(15) $\text{app}(\mathit{type}, \texttt{ATTR}) \sqsubseteq \downarrow$

(16) $\text{app}(\mathit{type}, \texttt{ATTR}) = \mathit{val}$

Since one may want to express that some properties are inappropriate for a certain class of objects, one can also assume that the value of the function **app** is undefined ($\uparrow$) for an attribute in a type.

(17) $\text{app}(\mathit{type}, \texttt{ATTR}) = \uparrow$

One possibility to implement these definitions would be to treat the $\downarrow$ and $\uparrow$ symbols as two types with an empty intersection and to make sure that all types other than $\uparrow$ are treated as subtypes of $\downarrow$.

On the basis of appropriateness definitions the description of an object by means of a typed feature structure can be constrained to be *well-typed*. Well-typed feature structures satisfy the conditions that every feature occuring in a typed feature structure is defined for the respective type and that the (typed) value of every feature is included in the type which has been defined as appropriate for that attribute.

(18) A feature structure of type t is well-typed if

$$\forall\ \texttt{F}_i,\ {}_{t}\begin{bmatrix} \vdots \\ \texttt{F}_i\ \mathit{val} \\ \vdots \end{bmatrix} \rightarrow \begin{array}{l} \text{app}(t,\ \texttt{F}_i\ ) \sqsubseteq \downarrow \quad \text{and} \\ \mathit{val} \sqsubseteq \text{app}(t,\ \texttt{F}_i\ ) \end{array}$$

This condition can be further restricted to define *totally well-typed* feature structures. While well-typedness only requires that every feature which occurs in a feature structure must be defined for the respective type, a totally well-typed feature structure of a given type must fulfill the condition that every feature which is appropriate for that type is also present in the feature structure.

(19) A feature structure of type t is totally well-typed if

a) it is well-typed and

b) $$\forall\ \texttt{F}_i\ ,\ \text{app}(t,\ \texttt{F}_i\ ) \sqsubseteq \downarrow \ \rightarrow\ {}_{t}\begin{bmatrix} \vdots \\ \texttt{F}_i\ [\cdots] \\ \vdots \end{bmatrix}$$

It should be clear that these conditions are restrictions over feature structures and not restrictions which have to be satisfied by a (partial) feature description. However, they can be employed to constrain and define operations over feature descriptions in various ways.

The notion of well-typed feature structures appears to be useful to constrain the necessary unification operations to term unification with types. Since the features defined as appropriate for a type are finite, they can be represented in a term with a fixed arity.

It is a highly debated issue, however, whether the latter notion of totally well-typed feature structures is a necessary or useful concept. This notion appears to play an important role only if *type inference* is used as a deduction method.

For example, in [Pollard and Sag 1994][21] it is defined that only feature structures which are *type-resolved* and totally well-typed are admitted as (total) models of linguistic entities, where type-resolved means that a feature structure has to be assigned the most specific type for which the features of the feature structure are defined as appropriate. This definition can be exploited to infer the well-formed feature structures for linguistic objects by means of type deduction. The simplest version of such an inference algorithm would roughly go as follows:

Given a feature structure of type t_j with attributes $\mathtt{F}_1$ ... $\mathtt{F}_i$. Find the most specific subtype t_k of t_j, for which the features $\mathtt{F}_i$ ... $\mathtt{F}_n$ are defined as appropriate.[6] Expand this type t_k to be totally well-typed. Descend and apply the algorithm to the type value of every feature $\mathtt{F}_1$ ... $\mathtt{F}_n$ which has been found for t_k.

Under the assumption of recursive type definitions, this algorithm will of course only terminate if it is constrained by some feature (path) in the input feature structure. The input could be, for example, a feature structure with an attribute representing the string or list of terminals to be analyzed, together with the (type) information that this structure represents some linguistic object. In this case, the syntactic, semantic or other properties would be inferred which have been defined as appropriate for linguistic objects and which are compatible with the decomposition of the string. Conversely, the input could be some semantic representation and the algorithm would have to find the string(s) appropriate for this structure.[7]

Definitions with Arbitrary Feature Constraints

The appropriateness definitions as we introduced them above have two severe restrictions. They are constrained to type definitions of feature structures of

[6] This most specific subtype may not be unique. The problem is also further complicated by multiple inheritance. See below.

[7] One of the most prominent examples of a system using this derivation method is the TFS system by Emele and Zajac.Cf. [Emele and Zajac 1990] and [Zajac 1992].

depth one, i.e., it is not possible to express directly that for a given type the value of a path of attributes has to be of a certain type. This can be done only indirectly by iteratively defining specific subtypes which eventually lead to the respective value. For example, it is not possible to define a type for nominative noun phrases such that only the top level type and the deeply embedded attribute for `CASE` are specified.

(20)
$$ {}_{\mathit{nom\text{-}np}}\left[\texttt{SYNSEM}\ {}_{\mathit{synsem}}\left[\texttt{LOCAL}\ {}_{\mathit{loc}}\left[\texttt{CAT}\ {}_{\mathit{cat}}\left[\texttt{HEAD}\ {}_{\mathit{head}}\left[\texttt{CASE}\ \mathit{nom}\right]\right]\right]\right]\right] $$

Instead, the embedded variation of the `CASE` value has to be averted at every level of embedding:

(21)
$$ {}_{\mathit{nom\text{-}np}}\left[\texttt{SYNSEM}\ {}_{\mathit{s\text{-}n\text{-}np}}\left[\texttt{LOCAL}\ {}_{\mathit{l\text{-}n\text{-}np}}\left[\texttt{CAT}\ {}_{\mathit{c\text{-}n\text{-}np}}\left[\texttt{HEAD}\ {}_{\mathit{h\text{-}n\text{-}np}}\left[\texttt{CASE}\ \mathit{nom}\right]\right]\right]\right]\right] $$

Second, the appropriateness definitions themselves do not provide the means to express that a type is defined for two attributes (or even two paths of attributes) which share the same value.[8]

To avoid these restrictions and to provide the basis of grammar specifications which do not require constraint definitions outside the domain of the type lattice, we follow the approach taken in $\mathcal{TDL}$ and allow types to be defined through arbitrary complex constraint sets. Thus a type definition will be simply of the form (22), where ϕ can be an arbitrary boolean expression over types and feature constraints.

(22) $\mathit{type}\ :=\ \phi$

This does not necessarily mean that we abandon the notion of appropriateness altogether. It simply means that type definitions of this form are given the interpretation that, if we know that an object is of type t, we know that the properties

[8] Possibly, it is one of the consequences of these restrictions that the sample grammar in [Pollard and Sag 1994][395ff] does not introduce any type definitions which go beyond feature structures of depth one. Any constraints referring to larger feature structures or introducing the sharing of some value are defined in the *principles*, which are not sorted into the type lattice but applied as constraints to given (typed) feature structures.

described in the constraint set are appropriate for an element of this type. Again, we will use abbreviatory graphical representations whenever possible, following the notation in (23).

(23) $$\textit{type} := \begin{bmatrix} \texttt{A} & {}_{t_1}[\;] \\ \texttt{B} & {}_{t_2}\begin{bmatrix}\texttt{C} & \boxed{1}\end{bmatrix} \\ \texttt{D} & \boxed{1}\,{}_{t_3}\begin{bmatrix}\texttt{E} & t_4\end{bmatrix} \\ \texttt{F} & *\textit{undef}* \end{bmatrix}$$

This type definition is given the interpretation that the attributes `A, B, D` are defined as appropriate for *type* and that the attribute `F` is undefined for that type. The values of `A, B, C` are constrained to the types t_1, t_2 and t_3, respectively. Any feature structure of that type must have token-identical values for the paths `B C` and D with the value being constrained to the type t_3. The value of the attribute `E` under the path `D` is constrained to the type t_4.

2.3.3 Type Hierarchies

As mentioned above, types can be interpreted as names for sets of objects. The relations between different types can thus be defined in terms of set relations, such as set inclusion, intersection etc. These relations will then define a partial ordering over types which can be represented as *type hierarchies*.

The simplest relation of set inclusion corresponds to a *supertype* – *subtype* relation between two types, such that if a type t_2 is a subtype of a type t_1 ($t_1 \geq t_2$) then the set of objects denoted by t_2 is a subset of the set of objects denoted by t_1.

Given that subtype relations correspond to set inclusion, it follows that an element in a subtype must have all the properties that are appropriate for the elements of the supertype (plus other properties which may be appropriate only for the subtype). On this basis an *inheritance* relation between a supertype and its subtypes can be defined.

This inheritance relation defines that if a type t_2 with $t_1 \geq t_2$ is assigned a feature structure ϕ, it holds that ϕ inherits all features which are defined as appropriate for t_1 and that the value of every feature $\texttt{F}_i$ in ϕ which is defined for t_1 must have an equally or more specific value in the type t_2. For type definitions which are specified by the appropriateness function above, this can be expressed as follows:

(24) Given two types t_1 and t_2, such that $t_1 \geq t_2$,

$$\forall\ \texttt{F}_i,\ \text{app}(t_1,\ \texttt{F}_i) \sqsubseteq\ \downarrow\ \ \rightarrow\ \ \begin{array}{l}\text{app}(t_2,\ \texttt{F}_i) \sqsubseteq\ \downarrow \text{ and} \\ \text{app}(t_2,\ \texttt{F}_i) \sqsubseteq \text{app}(t_1,\ \texttt{F}_i)\end{array}$$

Just as a set can be the subset of more than one set, a type can be a subtype of more than one supertype, which is given the interpretation that its denotation is in the intersection of the sets denoted by its supertypes. (It denotes exactly the intersection, if no additional feature constraints are added.) A subtype which stands in such a *multiple inheritance* relation inherits the union of all the properties which have been defined as appropriate for its supertypes. Again, for the simple appropriateness function this is defined as follows:

(25) Given a type t and a set of types $t_1, \ldots, t_n$, such that $t_1 \geq t, \ldots, t_n \geq t$,

$$\forall t_i,\ \mathtt{F}_j,\ \mathrm{app}(t_i,\ \mathtt{F}_j) \sqsubseteq {\downarrow} \quad \rightarrow \quad \begin{array}{l} \mathrm{app}(t,\ \mathtt{F}_j) = {\downarrow} \text{ and} \\ \mathrm{app}(t,\ \mathtt{F}_j) \sqsubseteq \mathrm{app}(t_i,\ \mathtt{F}_j) \end{array}$$

From these definitions it follows that two types must have an empty intersection, i.e., no common subtype other than the $\perp$ type, if they are defined by (or inherit) two conflicting constraint sets. This is the case, for example, if for the two types the same attribute is specified as defined and undefined respectively, or if there is some conflict of atomic values.

For the less restricted version of type definitions with arbitrary constraint sets, we will simply assume that the subtype inherits every feature path from the supertype and that the type of every feature value on this path is the most specific type specified either in the supertype or in the subtype. This means that if the specified type of a feature value on the subtype is less specific than the one introduced by the supertype, the more specific one will be chosen. While such a definition of a subtype would only contain some redundancy, we also envisage that the value of an attribute does not have to be typed at all.

To take our hypothetical example of an *np-nom* type definition from above, a type definition may introduce just a feature path without having to repeat type information on this path.

(26) $$np := sign\ \&\ \left[\mathtt{SYNSEM}\ {}_{synsem}\!\left[\mathtt{LOCAL}\ {}_{loc}\!\left[\mathtt{CAT}\ {}_{cat}\!\left[\mathtt{HEAD}\ head[\ ]\right]\right]\right]\right]$$

(27) $$np\text{-}nom := np\ \&\left[\mathtt{SYNSEM|LOCAL|CAT|HEAD|CASE}\ nom\right]$$

The general format for type definitions thus follows the pattern in (28). Type definitions which involve inheritance from one or more supertypes we write as the type name followed by a definition sign :=, a conjunction of the supertypes and an optional feature description defining the appropriate attributes and values for the type.

(28) *type* := *supertype*$_1$ & ... & *supertype*$_n$ & [ ]

For type definitions of this form we assume that the defined type can be sorted into the type hierarchy as the immediate subtype of its supertypes according to an *immediate subtype* relation (29).

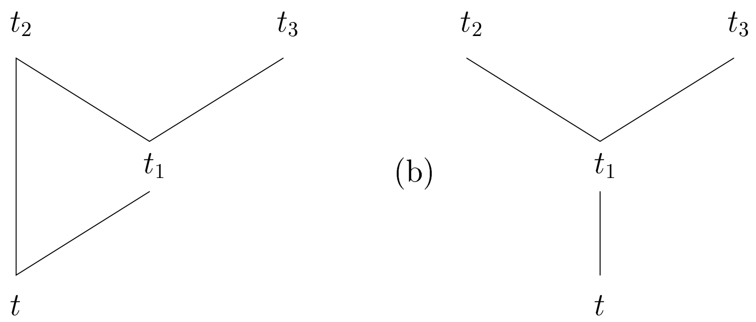
(29) A type t_2 is an immediate subtype of t_1 iff $t_1 \geq t_2$ and there is no type t_i such that $t_1 \geq t_i$ and $t_i \geq t_2$.

Thus, in a type definition with multiple supertypes the defined type represents a greatest lower bound (glb) for the supertypes occurring in the body of the definition. Immediate subtype relations can be read off the type definitions as long as there are no conflicts with the above definition. Such a conflict could occur for example through definitions such as (30).

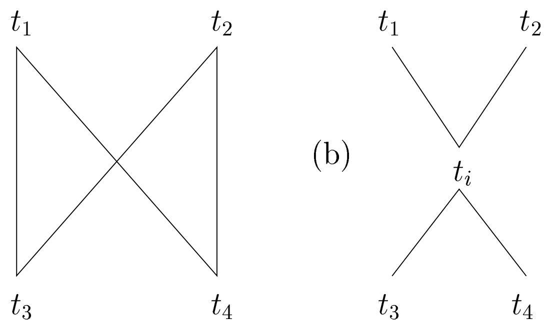
(30) t := t_1 & t_2

t_1 := t_2 & t_3

However, rather than ruling out such definitions, we will assume that the definition of the type t in (30) is not given the interpretation in (31a) if it occurs in the context of the definition for t_1, but rather the interpretation in (31b).

(31) (a)


Furthermore, we do not impose the constraint on type definitions that every set of types must have a unique glb (bounded completeness), but also allow definitions which would lead to configurations such as (32a). Whether or not these relations are resolved to the configuration in (32b), where a unique glb t_i for the two supertypes t_1 and t_2 is introduced, is left to the interpreter of the type system.

(32) (a)


2.3.4 Incompatibility of Types

As we already mentioned above the conjunction of two typed feature structures denotes an empty intersection ($\perp$), if the constraint sets associated with the two types do not unify.

However, we may also want to declare that two or more types are incompatible with each other and denote sets of objects with an empty intersection, even though the properties or features associated with the types do not lead to any obvious conflicts. The intuition behind this assumption is that we may want to classify the objects that we talk about into disjoint sets without having to assign to them dimensions of properties with conflicting values.

Of course, one way of achieving this objective is to declare that two disjoint types are respectively defined and undefined for certain attributes which we take to be appropriate or inappropriate for these types and which would lead to a conflict on the feature side. The possibility to declare types as incompatible simply gives us some more flexibility. Different techniques for declaring incompatibility are a closed world assumption, partition and local incompatibility declarations.

Open and Closed World Assumption

Most systems make a *closed world assumption* with respect to the type hierarchies. This means that for every conjunction of two or more types a unique greatest lower bound has to be defined. If no such glb is defined, the conjunction of the types denotes an empty intersection and the glb is given by the $\perp$ symbol. Together with the well-typedness constraint such a closed world assumption can lead to a quite efficient processing behaviour, since feature unification can be replaced by type unification, for which in principle linear complexity can be achieved.

The disadvantages of such a closed world assumption are on the other hand that it imposes rather strict constraints at specification time and thus impedes rapid proto-typing. To avoid these restrictions an *open world assumption* can be made, which can be complemented by different forms of incompatability declarations.

This simply means that the types at the yield of a type hierarchy do not automatically have the empty set as their glb. Thus, when defining the sets of objects, which one wants to talk about, one does not have to foresee every class of entities in the type definitions which may be formed by conjunction / intersection of two types. Nevertheless, the option is provided to forbid the formation of certain glbs by the declaration types to be incompatible. For compatible types the glbs can be computed either at compile-time or at run-time on the basis of the following definition.

(33) The conjunction of two types t_1 and t_2 with associated feature descriptions ϕ and ψ denotes a non-empty intersection of two sets of entities iff

a) the feature descriptions ϕ and ψ are satisfied by two feature structures F_ϕ and F_ψ denoting two sets with a non-empty intersection and

b) the two types t_1 and t_2 are not declared to denote an empty intersection of sets of entities.

Partitions

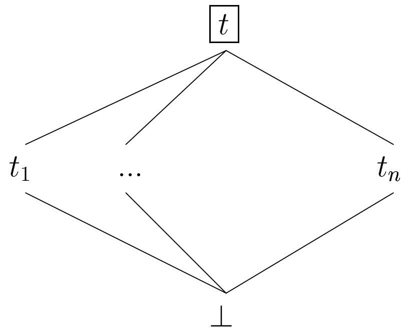
One other way to express the incompatibility relation between types is to declare a type to be entirely subdivided into disjoint subsets by means of *partition*. Closed world assumptions can be reduced to this relation, if all definitions of subtype relations are restricted to partition declarations, where all types which are not further partioned have $\perp$ as their glb. However, partitioning also makes sense in an open world assumption to define partial closure of the type hierarchy.

Partition is typically defined relative to a supertype and fully subdivides the set denoted by the supertype into disjoint subsets denoted by a given number of subtypes. Thus, if a set of objects denoted by a type t is partitioned by a number of subtypes $t_1, \ldots, t_n$, this means that the individual subtypes denote disjoint sets of objects and conversely that the union of the set of objects denoted by the subtypes is equivalent to the denotation of the supertype t. We will notate partition as a declaration **part** for a type which is assigned as its value a list of subtypes.

(34) $\text{part}(\mathit{type}) \quad = \quad \mathit{subtype}_1,\ \ldots,\ \mathit{subtype}_n$

In graphical illustrations we will mark partitions by boxing the partioned supertype (35). The $\perp$ symbol represents the empty set as a denotation, or failure of the conjunction of supertypes. For this symbol we have to assume a slightly different semantics in graphical representations than for types as glbs. If a type is depicted as the glb of a set of supertypes, this type denotes the intersection of all supertypes. For the $\perp$ symbol we assume that the intersection of any pair of immediate supertypes is the empty set, i.e., that the conjunction of any subset of the immediate supertypes leads to a failure.

(35)



Bottom Declarations

While partitioning means that *all* immediate subtypes of a type are incompatible and denote disjoint sets, we also allow the possibility to express that only *some* subtypes should have an empty intersection by means of *bottom declarations*. These declarations are even weaker than partitions, however, under an open world assumption they can help to express that some subtypes of a supertype are incompatible without enforcing the commitment that the denotation of the conjunction of all subtypes exhausts the denotation of the supertype. Thus, while partition represents a top-down perspective from a supertype to its subtype, bottom declarations are more like a bottom-up view from the $\perp$ symbol as a subtype to a number of supertypes.

We will notate bottom declarations as in (36), adopting again the convention that the types on the right hand side are pairwise incompatible.

(36) $\perp \;=\; type_1,\; \ldots,\; type_n$

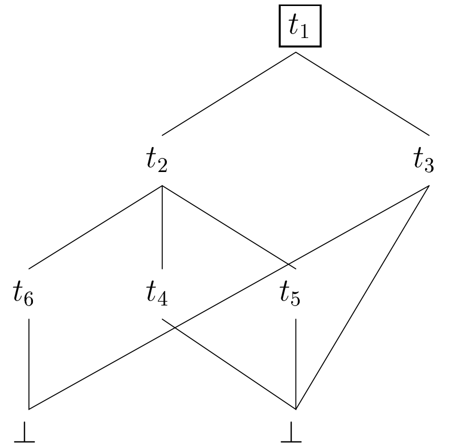
Bottom Propagation

Clearly, if two types are declared to be incompatible in some way or another, i.e., if they are said to have an empty set as the denotation of their intersection, this property will also be inherited by their respective subtypes.

In graphical representation this has as a consequence that we will have to introduce more than one $\perp$ symbol to avoid confusions. Take for example the set of type definitions and declarations in (37) which are graphically represented in (38).

(37) $t_2 \ := \ t_1$
$t_3 \ := \ t_1$
$t_4 \ := \ t_2$
$t_5 \ := \ t_2$
$t_6 \ := \ t_2$
$part(t_1) \ = \ t_2,\ t_3$
$\perp \ = \ t_4,\ t_5$

(38)



In (38) bottom propagation has the effect, that all types in the yield of the lattice are incompatible with each other except for the pairs t_4, t_6 and t_5, t_6. Since t_2 and t_3 are incompatible, all subtypes of t_2 are also incompatible with t_3. However, among these subtypes only t_4 and t_5 have been declared to be incompatible with each other, which means that individual combinations with t_6 are consistent.

These relations also constrain the definitions that can be added to (36). For example, the definitions in (36) can be augmented by the definitions in (39), while they are inconsistent with the additional definitions in (40).

(39) $t_7 \ := \ t_4 \ \& \ t_6$
$t_8 \ := \ t_2$

(40) $t_9 \ := \ t_1$
$t_{10} \ := \ t_7 \ \& \ t_3$

The first definition in (40) is inconsistent with the partitioning of t_1, which must not have any other immediate subtypes other than t_2 and t_3. The second fails again due to bottom propagation; since t_7 is an indirect subtype of t_2 it inherits the incompatibility with t_3 introduced by the partioning of t_1.

Types and Sorts

We introduced some of the additional machinery for declarations of incompatibility to allow for an open world assumption for the type hierarchies. However, we do not want this open world assumption to hold over the entire universe of our objects.

As mentioned above, we make a distinction between *types* and *sorts*. Types we described as names for classes of objects which may also be described by feature constraints. Sorts on the other hand we view as names for sets of unanalyzed objects for which no such additional description by feature constraints is made.

Remember that in choosing our notations we have deliberately selected a graphemic representation which leads to an ambiguity between complex types on the one hand and atomic values of attributes on the other. While constants or atomic values are assumed to be always incompatible with each other, it can be useful also to allow hierarchies over classes of atoms, which would correspond to sorts in our terminology. Atoms under this view are then instances of classes of singleton sets.

Since we do not enforce types to be associated with feature constraints, it makes sense to assume an open domain for types and a closed domain for sorts. This assumption will give us the full range of expressivity for describing different relations. Labels for classes of objects which are not associated with feature constraints can either occur in the closed domain of sorts, where we automatically infer $\perp$ if no glb is defined, or they can occur in an open world which may be partially closed by incompatibility declarations.[9]

To formalize this distinction one can partition the universe of objects ($\top$) into two classes *type* and *sort* where by convention the sublattice under *sort* is closed, whereas the one under *type* is not.

(41) $\text{part}(\top) \;=\; \textit{type}, \textit{sort}.$

Wherever this distinction is obvious in our type and sort definitions, below we will indulge in a certain sloppiness and simply assume that by convention the above assumptions hold and do not have to be spelt out in every case.

Types and Instances

A final distinction that we have to assume concerns types and *instances* formed from a type. Instances can be seen as those leaves of a type hierarchy, which

[9]An alternative to this approach, which so far has not been realized in $\mathcal{TDL}$ could be to dismiss the distinction between sorts and types and to allow partial sublattices of the type hierarchy to be declared as closed domains.

describe the individual objects in the universe. The idea is that we can introduce an arbitrary number of instances of the elements of a type, which are all assumed to be mutually disjoint. In this sense, instances of a type t could also be expressed as terminal, partitioning subtypes of t each describing a singleton set of objects.

For linguistic purposes this notion plays a role insofar as only a small set of types denote sets of objects occuring as linguistic *signs*. Thus, we can describe sets of lexical items as instances of a certain type, all of which may have individual properties and all of which describe disjoint elements. For example, we could assume a type for transitive verbs and then form specific instances for individual items such as *see*, *buy*, etc. Since there will be a multitude of types which only describe (sub-)properties of linguistics objects, the formation of instances also allows us to make explicit what objects make up our grammar in the sense of lexical and phrasal signs.

Since instances and types do not differ in their form of definition, we do not introduce a specific notation for instances. We simply assume that certain type definitions can be declared as definitions of instances, and accordingly will be interpreted as denoting singleton and disjoint sets and as having no further subtypes.

2.4 Cross-Classification and HPSG Principles

Since the notion of principles plays a very crucial role in HPSG, we will try to explore in this section by what formal means principles could be expressed in a typed feature formalism. To be more precise, we will discuss the ontological status of principles with respect to type hierarchies as representations of linguistic objects. In particular, our objective will be to show, how principles can be integrated in type hierarchies. As we will argue in more detail below, when we discuss the linguistic notion of universal principles in HPSG, principles under this view are more appropriately considered as generalizations in both the linguistic and formal sense.

2.4.1 Principles and Implicational Constraints

Principles in the earlier versions of HPSG are defined as implicational constraints following the pattern in (42).

(42) $\textit{antecedent}[\;] \rightarrow \textit{consequens}[\;]$

The interpretation is that if a feature structure has certain properties, defined in the (type of the) antecedent, then it has also the properties defined in the consequens. To put it differently a feature structure satisfies an implicational constraint if its denotation is either not contained in the denotation of the antecedent or if it is in the denotation of the intersection of the antecedent and the consequens.

One way to treat such implicational constraints is to view them as inference rules which are expressed independent of a type hierarchy. If a typed feature structure is found, which matches the type and the feature constraints in the antecedent, the constraints in the consequens can be inferred and unified with this feature structure. Such an inference rule could make sense as a feature co-occurence restriction, where from one property the presence of another property is inferred.

It is interesting to note however that as far as we can see principles in HPSG are mostly of a kind where the properties defined in the antecedent are also present in the consequens. One typical example is the Head Feature Principle, which says that in a feature structure of the type *headed-structure*, i.e., a structure which contains a feature HEAD-DTR, the HEAD value under this attribute is token-identical with that of the mother.

(43) $\textit{headed-structure}[\;] \rightarrow \begin{bmatrix} \texttt{SYSNEM} & \begin{bmatrix} \texttt{LOCAL|HEAD}\ \boxed{1} \end{bmatrix} \\ \texttt{HEAD-DTR} & \begin{bmatrix} \texttt{SYNSEM|LOCAL|HEAD}\ \boxed{1} \end{bmatrix} \end{bmatrix}$

Now, the question is what status the consequens in this definition has with respect to a type hierarchy into which the type *headed-structure* is sorted. One possible answer is "none", since it is not typed at all. However, if one wanted to assume that this structure should be typed, since there should be no structures which do not carry a type label, we are in a dilemma. Either we assume that the type of the consequens is *headed-structure*, in which case the principle is superfluous, since the "inferred" information would be contained in the definition of this type. Or we assume that it has a type other than *headed-structure*. In this case there are basically three possibilities.

(a) The type is a supertype of *headed-structure*. In this case the appropriateness definition of *headed-structure* which only introduces the feature `HEAD-DTR` is practically redundant, since this feature is already contained in the supertype.

(b) The type is a subtype of *headed-structure*. In this case it may be odd to express this relation as an implicational constraint, since now the denotation of the antecedent is a superset of the denotation of the consequens, i.e., the intersection of the antecedent and the consequens is trivially reduced to set inclusion.

(c) The two types are not in a supertype–subtype relation at all. Again this would be either odd, since either both types introduce the attribute `HEAD-DTR` as appropriate and there would have to be a common subtype as their intersection, or there is a common supertype introducing this feature, which makes the definition of *headed-structure* redundant again.

For our purposes we can concentrate on the first two options, since (c) can be reduced to either (a) or (b) since the introduction of a respective additional supertype or subtype is practically equivalent to one of these two versions.

As mentioned option (a) makes the definition of the type *headed-structure* redundant. The only justification for such a type would be to combine the information of different principles which have this type in their antecedent by means of multiple inheritance. However, in terms of information structure this strongly implies that there should be a supertype of the consequens which contains less than this information.

As to (b), this version takes into account the constraint sets referred to in the principle, capturing the fact that the set of structures which are defined for the attribute `HEAD-DTR` is a superset of those structures which have in addition the property of coreferent `HEAD` values. However, the purpose of the HFP is to express that there must not be any (well-formed) structures which have the first property but not the second. If we want to account for this aspect of the principle, we have to conclude that the sets of objects denoted by the types corresponding to the antecedent and the consequens of the principle must be coextensive.

Of course, one way to capture this latter relation would be to interpret principles like the HFP simply as type definitions, where the antecedent is the defined type. However, since different principles may have the same antecedent, we would have to allow the same type to be defined more than once or we would have to conflate different principles into one type definition, which would go against the modularity which can be achieved through type inheritance.

In particular, one very central assumption in HPSG is that every well-formed (phrasal) sign must satisfy the definition of exactly one rule schema plus the conjunction of all principles. This assumption could be described by representing all principles as types in the type hierarchy and to intersect them with the rule schemata.

The question is, however, whether we can express the conjunctive restriction on the relation between principles and rule schemata (or phrasal signs in general) as a necessary part of our definitions. I.e., how can we prevent the definition of a phrasal sign, which obeys only a subset of the principles and maybe none of the rule schemata?

While in the formulation of principles as (separate) implicational constraints this is 'guaranteed' by testing a given sign for subsumption of the antecedents of all principles, this is not a priori given if principles are represented as types of the type hierarchy.

2.4.2 Cross-Classification

In the following, we explore an alternative approach which takes into account that principles are integrated parts of a type hierarchy, but also captures the implicational force of principles. For this purpose we introduce an extension to type declarations which can be seen as the dual to partition.

Whereas partition defines that the intersection of the subtypes of a type is empty, we want to be able to express that an object of a certain type can be subdivided into two or more subtypes which must have a non-empty intersection.

A weak version of this constraint would be a well-formedness constraint on the type definitions to the effect that the glb of two (or more) specific subsorts must be defined and that it must not be the empty set $\perp$. We could notate this by inequality of $\perp$ and a list of types, adopting the abbreviatory convention that the intersection of any pair of types must not be empty and that the intersection of the entire list of types must not be empty.

(44) $\perp \;\neq\; type_1,\; \ldots,\; type_n$

What this would buy us is the guarantee that two subtypes of a type must not be contradictory, and that a type hierarchy is well-formed only if a common subtype exists or can be inferred.

For our purposes such a relation would be too weak and has to be strengthened to a relation, which says that every element contained in the supertype must also be an element of each of the subtypes. Taking up a term used in [Pollard and Sag 1987][198ff], we call this relation *cross-classification* (**cross**) with the following syntax and semantics.

(45) $\text{cross}(type) \;\equiv\; subtype_1, \ldots, subtype_n$

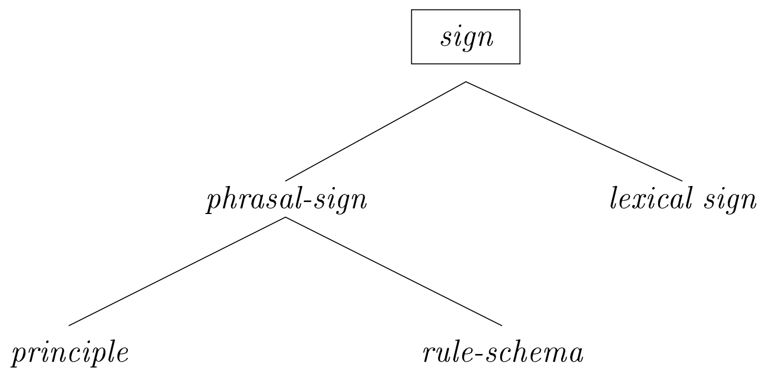
A type t is cross-classified by a set of subtypes $t_1, \ldots, t_n$, iff every element which is an element of the t is also an element of each of the subtypes.

We illustrate this relation by a type hierarchy which roughly corresponds to the HPSG assumptions about the organization of the top strata of linguistic objects (46). To begin, we partition the set of signs into phrases and words. Assuming for the moment that principles only apply to phrasal signs, we introduce two subsets of the type *phrasal-sign*, *principle* and *rule-schema*.

(46)
$$\begin{array}{lcl} \textit{phrasal-sign} & := & \textit{sign} \\ \textit{lexical-sign} & := & \textit{sign} \\ \textit{principle} & := & \textit{phrasal-sign} \\ \textit{rule-schema} & := & \textit{phrasal-sign} \\ \text{part}(\textit{sign}) & = & \textit{phrasal-sign}, \textit{lexical-sign} \end{array}$$

These definitions correspond to the following type hierarchy, in which we signal partition by boxing the respective supertype.

(47)



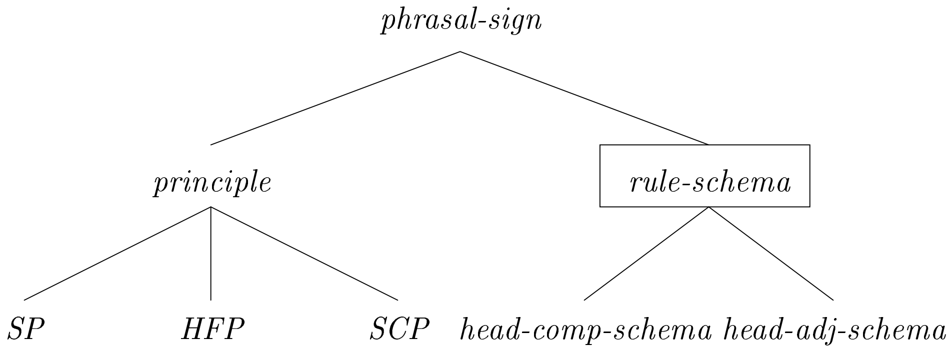
For the sake of illustration we only assume three principles, the Head Feature Principle *HFP*, the Subcategorization Principle *SCP* and the Semantic Principle *SP*.

$$
(48) \quad \begin{array}{lcl} \mathit{HFP} & := & \mathit{principle} \\ \mathit{SCP} & := & \mathit{principle} \\ \mathit{SP} & := & \mathit{principle} \end{array}
$$

The type *rule-schema* we assume to be partitioned by only two subtypes at the moment, the *head-comp-schema* for complement structures and the *head-adj-schema* for adjunct structures. Partitioning will already capture the fact that every phrasal sign must not satisfy two rule-schemata simultaneously.

$$
(49) \quad \begin{array}{lcl} \mathit{head\text{-}comp\text{-}schema} & := & \mathit{rule\text{-}schema} \\ \mathit{head\text{-}adj\text{-}schema} & := & \mathit{rule\text{-}schema} \\ \mathrm{part}(\mathit{rule\text{-}schema}) & = & \mathit{head\text{-}comp\text{-}schema}, \mathit{head\text{-}adj\text{-}schema} \end{array}
$$

(50)



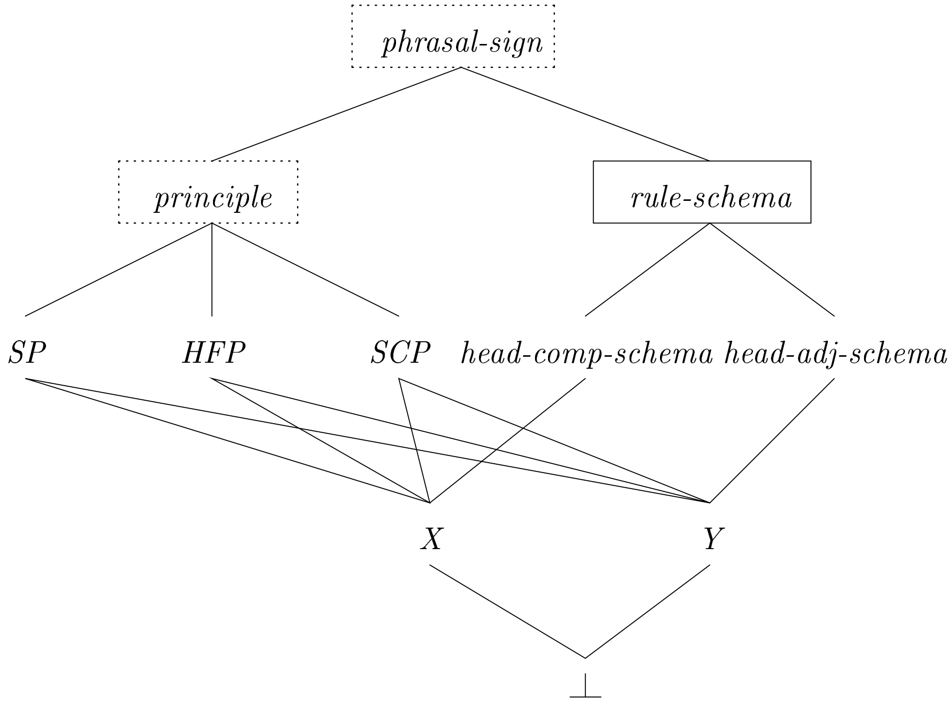
To capture the assumption that every phrasal sign must obey all principles and be an element of a rule-schema, it does not suffice to declare that *principle* and *rule-schema* must have a non-empty intersection. This condition would already be satisfied, if some rule schema obeyed some principle.

What is required here is a cross-classification of the two types *phrasal-sign* and *principle*.

$$
(51) \quad \begin{array}{lcl} \mathrm{cross}(\mathit{phrasal\text{-}sign}) & \equiv & \mathit{phrasal\text{-}principle}, \mathit{rule\text{-}schema} \\ \mathrm{cross}(\mathit{phrasal\text{-}principle}) & \equiv & \mathit{HFP}, \mathit{SCP}, \mathit{SP} \end{array}
$$

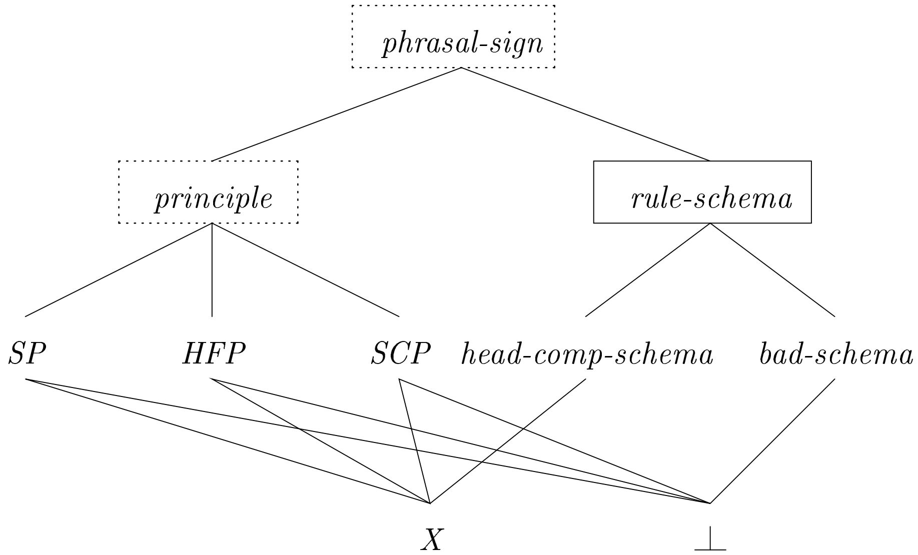
If we augment the above definitions by the declarations in (51), every phrasal sign must simultaneously obey all principles and it must be an element of the class denoted by *rule-schema*. Since this type is partitioned, the cross-classification is propagated further down, similar to the bottom propagation in partitions. We can infer the type hierarchy in (52), where cross-classification of types is represented by dotted boxes. In this lattice two 'dummy' types X and Y are introduced as the appropriate intersections between the principles and schemata. To be precise, this type lattice is the smallest lattice which satisfies the above definitions and declarations.

(52)



An interesting question is what happens if, say, a rule schema is defined such that it does not obey some principle. For example, we could define a schema with a head-daughter which has a `HEAD` value conflicting with the mothers `HEAD` value, thus violating the HFP. Now, the theory would predict that there must not be any phrasal sign of this kind. One way to represent this case would be (53), where the intersection of the schema with the conflicting values *bad-schema* with the principles would be the empty set.

(53)



However, one could even go further and reject the definition of such a type right away. The condition of cross-classifying phrasal signs into principles and rule-schemata implies that if an object is a phrasal sign it must be element of both sub-types. If the intersection of the a subtype of *rule-schema* with the principles yields an empty intersection, this can have two reasons. Either the principles are inconsistent, in which case the entire set of phrasal signs would be the empty set. If this is not the case, it must be that the respective subtype of the rule-schemata itself denotes an empty set, since only then its intersection with a non-empty set is empty. Of course one could allow the introduction of such a class inferring that its denotation is empty. However, since there is an infinite number of subclasses to a type which have an empty intersection, it makes just as well sense to reject the introduction of such a type altogether. Thus cross-classification would capture not only the necessary combinatorics of principles and other objects, it would also provide a filter over what objects can be defined at compile-time.

Chapter 3

Theory of HPSG

3.1 Overview

In this chapter we discuss some of the theoretical foundations of HPSG. We will address questions of the grammar model, specifically the issue of universal and language particular grammar, as well as the notion of a *sign*, representing the relation between the form and the content of a linguistic expression. We discuss the notion of *head* in HPSG, which plays a central role both in terms of projections from the lexicon and in terms of establishing grammatical relations between sibling constituents. Building on this discussion of grammatical relations, we suggest an alternative set of phrase structure schemata which are essentially binary branching and which discard with basically unmotivated unary structural projections. We show that this alternative approach can be exploited to express parametrization of word order and configurationality in a more adequate way. Finally, we show how principles and rule schemata can be represented in a typed feature formalism.

3.2 Grammar Model of HPSG

3.2.1 Universal and Language Particular Grammar

One of the expressed aims of the theory of HPSG, as it is formulated in both [Pollard and Sag 1987] and [Pollard and Sag 1994], is to describe the relation between a putative *Universal Grammar* and *Language Particular Grammar*. This relation builds on a threefold distinction between principles, phrasal signs and lexical signs. Principles are basically generalizations over different representations of linguistic signs, phrasal signs refer to configurations and relations between linguistic signs, and lexical signs model atomic linguistic expressions, which are internally only structured on the levels of morphology or phonology.

In [Pollard and Sag 1987] the relation between universal and language particular grammar is reduced to the formulae (54) and (55):

(54) **Universal Grammar:**

$$\mathrm{UG} = P_1 \wedge \ldots \wedge P_n$$

(55) **Language Particular Grammars:**

$$\text{English} = P_1 \wedge \ldots \wedge P_{n+1} \ldots \wedge P_{n+m} \wedge (L_1 \vee \ldots \vee L_p \vee R_1 \vee \ldots \vee R_q)$$

where $P_{1 \ldots n}$ represent a set of universal principles, $P_{n+1 \ldots n+m}$ a set of language particular principles, $L_{1 \ldots p}$ a set of language-specific lexical signs and $R_{1 \ldots q}$ a set of language particular rule schemata. The rule schemata are typically assumed to be representations of immediate dominance relations with linear precedence constraints coming in through language-specific LP principles.

The intuition behind these formulae is that every language particular grammar has to obey a set of universal principles as well as some language-specific principles. The set of these principles has to be obeyed by every expression in a language, which must be either a lexical or a phrasal sign.

In [Pollard and Sag 1994][56] this definition is further refined and strengthened.

(56) **Universal Grammar:**

Linguistic Ontology: the inventory of universally available sorts of linguistic entities, together with a specification of their appropriate attributes and their value sorts.

Schemata: a small, fixed inventory of universally available phrase types (schematic immediate dominance schemata) [...]

Universal constraints on well-formed phrases: head feature principle, subcategorization principle, etc.

(57) **Particular Grammar:**

Lexicon: a system of lexical entries (possibly interrelated with lexical rules).

Linguistic Ontology: selection from and further articulation of the linguistic ontology.

Schemata: selection from and further specification (e.g., for the constituent order) of universally available schemata.

The linguistic ontology we can assume to be represented in a system of type definitions or typed feature structures. The fact that this characterization does not mention language-particular principles anymore presumably has to be seen as an idealization. The elimination of such principles in favour of parametrized universal principles is mentioned in [Pollard and Sag 1994][20] only as a long-term goal. An interesting question which we do not have to address here is whether this setup is excluding generalizations over the lexicon in different languages. For example, it would be hard to express in this setup that lexical items normally do not select more than one subject or more than one direct object. While such generalizations form an inherent part of theories, such as LFG, they would be hard to express if no generalization over lexical structures are allowed, where selection typically resides in strongly lexicalized theories.

The most crucial change, however, appears to be that this revision licences language-particular rule schemata only if they can be reduced to a universal rule schema, i.e., language-particular rules have to be contained in the set of universally available rules. We will discuss this assumption together with the question of the status of universal principles in the next section.

3.2.2 Predictive Power of Universal Principles

It is quite obvious that the distinctions made above are more likely to be interpreted as a research program rather than an implemented and verified theory. However, the question is whether we can derive any more precise criteria which allow us to determine whether a specific theory is consistent with this grammar model. What appears to be orthogonal to this issue is that we can always take a deductive or an inductive approach (or any mixture of the two). We could start with postulating a set of principles as universal and then attempt to falsify and refine these assumptions on the basis of language-specific data. Alternatively, we could start with any language, for example English, and then derive from this language universal generalizations by abstracting from what appears to be too language-specific.

However, in both cases we need some criteria which tell us what permissible and necessary abstractions are. To put it crudely: if we see a universal generalization, how do we know it.

Universal Rule Schemata

To approach this question for rule schemata, we can first ask how the set of universal schemata could be defined more precisely. The way it is specified in (56) this set of schemata $\mathcal{S}_u$ corresponds to disjoint sets of descriptions of phrasal signs $\mathcal{S}_{u_1}$ ... $\mathcal{S}_{u_n}$, each set obeying certain constraints. Each language-particular schema has to describe a subset of one of these sets, i.e., $\mathcal{S}_{p_m} \subset \mathcal{S}_{u_m}$.

In a deductive approach, the application of this criterion appears to be straightforward at first sight. We start with a given set of universal schemata. When we describe a language-particular phenomenon, this has to be described by a rule schema which is a more specific instance of a universal schema. If a phenomena cannot be subsumed under such a universal schema, we have falsified our initial assumption and we have two choices. Either we generalize one of the universal schemata until it allows the subsumption of the language-specific phenomenon, or we add another schema to the universal set, which obeys this subsumption relation.

In an inductive approach, one could start with an extreme position and define the set of universal schemata as the disjunction over all disjoint sets of language-particular schemata. Such an assumption would still be consistent with the characterization given above, however in a trivial way, since one would simply equate the set of all phrasal schemata in all languages as the set of universal schemata. Since it is claimed that the set of universal schemata should be "small", one could of course argue that those characterizations which contain the smallest set of disjoint sets of descriptions of schemata are the most adequate ones. To achieve this objective of reducing the set of disjoint schemata, generalization in the formal sense would be the correct approach.

However, here the crucial issue emerges, which holds for both the deductive and inductive approach. The above definition could be satisfied equally trivially, if the set of universal rules schemata were (next to) empty or if the constraints imposed by this set reduced to statements of practically non-commitant properties. Thus, if we apply generalization to a set of disjoint schemata, when do we stop generalizing? Do we have a criterion, which stops us from generalizing too far?

The point is that, as long as languages can draw arbitrarily on the set of universally available schemata, we loose any criterion for deciding on when a schema should make its way into universal grammar and when it should be considered as a language-particular instance of a more general universal type. Whether two schemata realized in one or more particular language should be considered as two sub-types of *one* universal schema or as separate subtypes of *two* universal schemata, appears to be undecidable as long as we do not impose any further constraints or add some other criteria.

The strongest, but not necessarily ideal form of such a criterion would be that every language-particular grammar must non-trivially include at least one instance from each language universal schema. 'Non-trivially' means that it does not suffice to add constraints to an instance of this schema until it can never be applied anyway, but rather that this instance has to be consistent with the rest of the grammar and lexicon, such that it can be applied to some well-formed constructions. Unfortunately, it is easy to demonstrate that this criterion is quite likely to turn out to be too strong.

Assume, for example, that by careful consideration of a number of languages we have come to the conclusion that for the description of non-local dependencies a *filler-head-schema* is required in these languages. Since no other schema subsumes this relation it is stipulated as a universal schema.

If we now come across a language which does not have long-distance dependencies of the relevant kind, we are in a dilemma. On the one hand the strong criterion that every language has to instantiate every universal rule schema, is not fulfilled by a possibly exceptional language. On the other hand if we attempted to satisfy the criterion, we would loose the universal schema which was originally postulated to account for treatment of the phenomena in a number of languages.

A logical consequence would be to generalize this schema further and reduce it to a simple *head-non-head-schema* which would subsume the *filler-head-schema* but which would accidentally subsume also all other schemata. Thus, such an approach could easily take us too far and lead to a theory of universal schemata which is almost void of any content. The only prediction it could make is that every language particular schema has to be a subset of this one single universal schema. This would practically take us "back" to categorial grammar, where universal grammar in some sense boils down to stipulating a general functor-argument relation.

Thus, if the notion of universal grammar is to play a non-trivial role in the grammar model and research program of HPSG, what appears to be required is a stronger theoretical commitment than just claiming that there is a set of universal schemata which are somehow instantiated in the different languages. The class of universal schemata has to be given some positive and substantial content, i.e., it has to be specified what aspects of language the universal schemata are supposed to capture.

One such possibility to fill the notion with content is to claim that universal schemata specify a limited number of abstract grammatical relations. These grammatical relations would be reflected, for example, in the labelling of daughters attributes in phrasal configurations and in the selection relations holding between these daughters.

While such a move clearly wouldn't take HPSG all the way through to LFG, it appears to entail that the quest for universal grammar should take a shift from the specification of phrase structure configurations more or less along the lines of $\overline{\mathrm{X}}$-theory to a more general definition of universally available very abstract grammatical relations. It will be a central concern of sections further down in this chapter to show how such a notion of universal schemata could be instantiated.

Universal Principles

Basically the same problem as for universal schemata occurs in the case of so-called universal principles as they are characterized above. At first sight, the model appears more constrained since it does not allow any language-particular principles anymore. The model specifies that for any (universal or language-particular) phrasal sign it must hold that it satisfies exactly one of the universal schemata and all of the universal principles.

However, if we look at the way principles are defined in HPSG this constraint does not impose any restrictions whatsoever. As already outlined in the previous chapter principles are typically defined as implicational constraints over typed feature structures following an abstract schema.

(58) "If a feature structure is of type *type-1* then it also has all the properties of a type *type-2*"

$$\textit{type-1}\big[\;\big] \;\Rightarrow\; \textit{type-2}\big[\;\big]$$

For a principle to 'apply' to a given phrasal sign the constraint specification of the antecedent of the principle has to be subsumed by this phrasal sign. Out of a conjunction of principles of implicational constraints normally only a small subset of the antecedents fulfills the subsumption constraint and thus only a small number of principles applies.

Given these general assumptions, the following situation may arise: Suppose, that for the description of some phenomenon in a particular language a principle is stipulated. It may then be the case that the constraints imposed by the antecedent of the implication are so idiosyncratic that they are not satisfied by any structure in any other language because this specific configuration of constraints only occurs in the one language considered first. Yet, this idiosyncratic principle can safely make its way into universal grammar, since it will never apply except in the one language it was designed for and thus the general constraint on universal principles is obeyed. Again, what appears to be missing is a general criterion which would block this principle to qualify as a universal one.

Of course, the less specific the application conditions (the antecedent of the implicational constraint) of a principle are, the wider is its application domain across different languages. Again, what we would expect in *the* universal grammar, are principles which find an application in every language or which are of such a nature that they will not lead to contradictions or incorrect predictions in different languages. Linear precedence constraints, which define, for example, the ordering of heads relative to complements, are unlikely to turn out to be universal, since we typically find clusters of languages following certain patterns, but hardly any constraint which is fulfilled by all languages.

Thus, if we attempted to operationalize the procedure of finding the set of universal principles by submitting them to empirical scrutiny, we may well end up with a strange mixture of so-called universal principles, comprising very abstract, presumably truly universally applicable principles and very idiosyncratic principles which can be applied to only some languages. Principles which may hold for large classes of families, on the other hand, are rightfully banned from universal grammar; however, given the simple dichotomy of universal and language-particular grammar they have to go straight into the latter.

3.2.3 Summary

What is the upshot of this discussion? a) We seriously doubt whether the characterization of universal grammar given above will allow us to constrain the set of universal rules and principles on a non-ad-hoc basis. b) Even if it may weaken the notion of a universal grammar as a model of some innate language capacity, we argue for a stratified layering of different degrees of universality, which takes into account generalizations across classes of languages. c) For the analysis of some fragment of German outlined below, we will accept only a small set of very abstract universal principles and rule schemata. What they are we will discuss in the next few sections. Conversely, we do not attempt to claim any universality for the analysis given, unless it could be corroborated by more extensive comparisons across languages. This does not mean that we try to dodge the issue of universality, but simply that we will already have made some progress if we manage to contrive a consistent analysis of some fragment of German, which obeys certain general constraints and which stands a chance of being generalizable across some languages.

3.3 On the Notion of 'Head'

In most theories of grammar, above all those which are considered to have a high degree of lexicalization, the notion of *head* plays a central role in the definition of structural configurations.[1] This notion is usually associated with two other types of grammatical relations, viz., the notion of phrasal projections from the lexicon, and the characterization of different grammatical relations within a phrase. On the one hand, we have the perspective under which heads are seen relative to the immediately dominating phrasal constituent, which also comprises the assumption that some daughter of a phrasal node is designated more or less explicitly as its head daughter and that certain properties of this daughter determine properties of the mother node or phrasal projection. On the other hand we have the sibling relations through which grammatical relations such as *complement-of*, *adjunct-of* or *specifier-of* are defined relative to a head.

3.3.1 Heads and Phrasal Projections

A very abstract motivation for the first concept, the *head-of-a-phrase* relation, lies in the assumption that for (almost) every phrasal constituent there is some daughter with a specific set of properties which are relevant also at the level of the mother node.

Historically, this usage of the term *head* can be traced back to early structural linguistics, as for example in [Bloomfield 1933][194ff], who distinguishes endocentric and exocentric constructions on this basis. Exocentric constructions are those in which the resulting phrase does not belong to the same class or category as one of its immediate constituents. Endocentric constructions, on the other hand, he defines as those phrases which belong to the same 'formclass' as one of their constituents, the 'head'.[2]

[1] For a brief overview on the different usages of the term *head*, see the bibliographical note in [Matthews 1981][163ff].

[2] In this context Bloomfield even defines a notion of lexical projection which is not much different from this notion in more recent frameworks:

> If all the syntactic constructions which go to make up a phrase a endocentric, then the phrase will contain among its ultimate constituents some word ... whose form-class is the same as that of the phrase. This word we will call the *center* of the phrase.

It should be noted that the distinction between endocentric and exocentric constructions is not always consistent across different usages. Whereas a more narrow definition is based on the criterion of substitutability, i.e., endocentric constructions are those in which one constituent can substitute the entire construction, exocentric constructions are those for which no such constituent can be found, a more relaxed usage of endocentric will also allow substitution of the

This notion finds its reflex in more modern formalizations of linguistic generalizations, as for example in $\overline{\text{X}}$-theory [Jackendoff 1977][29f], where it serves the function of constraining the admissible rules of universal grammar. The term *head*, or rather *head of a phrase* X^n, refers either to the category X^{n-1} that is dominated by X^n or to the lexical category X at the bottom of such a configuration. Jackendoff was also among the first to describe phrase structure in terms of features and who thus introduced an early notion of *head features*.

This notion, which is very closely related to the relation *head-of-a-phrase*, becomes increasingly relevant in those theories which make it their central concern to express equivalence classes or identity of properties between heads and their projections by means of attribute-value configurations. In principle, it matters very little whether these head features define the category (or part of speech) of a constituent, or whether they comprise, say, agreement features. Most theories differ only with respect to the implicit or explicit declaration of heads, the specific sets of properties which are being shared between the head of a phrase and the phrasal node itself, as well as in the degree of 'persistency' which holds over the sharing of features across different projections. We illustrate some of these differences and their consequences by comparing some aspects of the notion of head in HPSG with the ones developed in LFG and GPSG.

The Notion of Head in HPSG, LFG and GPSG

As to the declaration of heads, it could be argued that Lexical Functional Grammar [Bresnan 1982a][296ff], for example, does not foresee an inherently explicit declaration of the head of a phrase in all local configurations. Heads are defined as those phrase structure categories which bear an annotation which equates the daughter's f-structure with its mother's f-structure. However, since so-called minor categories may carry the same kind of annotation as a major category functioning as the head, the head of a phrasal constituent is not always inherently and unambiguously identified in a local tree (of depth one). The only pertinent information about the head is either expressed implicitly through the observation of some encoding prescription requiring a certain kind of tree node labelling involving some $\overline{\text{X}}$-vocabulary, or it has to be derived indirectly from the fact that only one daughter must eventually contribute a feature representing the predicate (or 'semantic form') of the phrase. Since the labelling of tree nodes in LFG is completely arbitrary (from a formal point of view), it is only this latter property which could be considered as an inherent criterion for the definition of head in LFG.

more general category. This terminological difference becomes relevant insofar as a VP with a transitive verb and an object is not an endocentric construction under the former reading, whereas it is endocentric under the latter.

Leaving aside this minor flaw, it has to be noted that LFG overcomes the distinction between endocentric and exocentric constructions on the level of phrase structure, by defining the notion of head on the level of functional structure. For example, LFG follows the common assumption that the category S is not a projection of the category VP, i.e., an exocentric construction in phrase structure terms, whereas the functional structure associated with this maximal projection is identified with the functional structure of the lexical head of the VP.

In contrast to the definition in LFG, which has to take into account both functional annotations, phrase structure category and potentially also the annotations to the lexical anchor, HPSG chooses to designate the head in a local configuration by an explicit labelling of one of the daughters as a HEAD-DTR. This entails, to a certain degree, that the function of head cannot be split across several daughters (as it is the case with minor and major categories carrying the same functional annotation in LFG) but has to be assigned uniquely to a single daughter.[3] Of course, this label is a priori a label as any others, however, it is through this designated label, referred to in various principles, that the essential head properties are constituted. It is almost a natural consequence that in HPSG the distinction between endocentric and exocentric constructions hardly plays a role anymore, as the only configurations which do not involve a head daughter are coordinative constructions, which in some approaches are also treated as headed phrases.

The second difference mentioned refers to the set of properties which are treated as *head features*. Again, in LFG there is hardly any distinction between different sets of attributes. Instead all properties associated with a mother node are identified with the properties of the node functioning as its head in the functional structure. This also means that the subcategorization requirements of a lexical head and all its phrasal projections are identical. The only information which is not shared between a head and its mother is again the information encoded in the node labels of the constituent structure.

In HPSG, on the other hand, only a subset of the properties of a HEAD-DTR are selected and 'bundled' as head features under a designated attribute HEAD. These features are subject to the locality restrictions expressed by the Head Feature Principle (HFP), which determines that only these features are being shared between a HEAD-DTR and its mother:

(59) In a headed phrase structure the HEAD features of the mother of the constituent are token identical with the HEAD features of the HEAD-DTR.

In some trivial sense this selective percolation of properties already involves a kind of 'pseudo-monotonicity' since some attributes of a head daughter may take

[3] In all fairness, it has to be mentioned that a similar situation as in LFG could arise in HPSG, if the string or value of the PHON feature corresponding to a HEAD-DTR were discontinuous.

a different value on its phrasal mother. However, there is also the strictly formal kind of non-monotonicity which we find, for example, in GPSG, where a subset of properties of a category can be declared as head features, but where it is also permissible to change the value of individual head features across phrasal projections. Since the Head Feature Convention in GPSG has the status of a default principle it is possible that a phrasal node does not share *any* of its properties with its head node; i.e., this theory admits that basically any category can be re-encoded into any other category.[4]

Again, HPSG is more constrained in this respect since it requires full token identity of head features across a phrasal projection, encoded in the Head Feature Principle. As a consequence of this principle, none of the 'pseudo-monotonic' or non-persistent properties, such as the degree of saturation of subcategorization requirements (completeness), information about the lexical or phrasal status of a constituent, or certain kinds of non-local dependencies, must be considered a `HEAD` property, but have to be separated out and handled by different mechanisms. Thus, the strictness of the assumption of token identity already entails some negative constraint on what properties of a phrase may be considered as head features.

Of course, one could argue here that the example of LFG, where all properties associated with a head node are also associated with its mother, shows that persistency in the percolation of head features does not impose any constraints whatsoever. However, it has to be kept in mind that in LFG there is also a substantial amount of non-persistent information which is not involved in this identity relation. We already mentioned that some non-persistent categorial (or semantic) information is encoded in different strata or levels (e.g., c-structure or semantic structure). However, there is also some non-monotonicity of a different kind involved. Subcategorization requirements in LFG are controlled by the conditions of coherence and completeness. At least the completeness condition can be argued to be non-monotonic also in the formal sense, since a violation of this condition can be checked only at the end of a derivation and may then lead to the failure of a so far consistent constraint set.

Restrictions on Head Features

As we just tried to show the notion of a head in HPSG is defined in a rather strict way in comparison to other theories. Principled assumptions such as the HFP entail certain restrictions on the choice of heads and in particular of head

[4] Cf. [Gazdar et al. 1985] and [Pullum 1991] as an illustration of this case. However, as Kornai and Pullum [Kornai and Pullum 1990][39ff] point out, this apparent lack of constraints has very little to say, since the same re-encoding of categories could clearly also be achieved in a framework obeying a monotonic head feature principle if one allowed the usage of empty heads.

features. However, while the strictness of the HFP forces us to factor out some features as head features, the question arises whether it is possible to define some positive heuristics to prevent an unnecessary proliferation of head features. While HPSG makes it more or less clear, what features *must not* be projected as head features, it is not quite as obvious, what features *must* be head features, and what features *need not* be head features. In short, the question is whether it is possible to give an independent, linguistically motivated characterisation of head features. To put it differently; is it possible to make any universal claims or assumptions which would allow us to partition features into a set of head features and a set of non-head-features, with possibly a small set of features in the intersection whose head feature status would depend on language-specific properties?

As already mentioned, one of the fundamental motivations behind the concept of head features is that it allows properties to be shared or projected across a dominance path which is defined by the transitive relation head-of-a-phrase. An important consequence of this fact is that some relations which are actually 'non-local' can be broken down into a chain of *strictly local* relations, where strict locality is meant to refer to domains of trees of depth one.

Take the simple example of subject–verb–agreement in the phrase

(60) $[_{NP}$ The little boys $]$ $[_{VP}$ often play with the big girls $]$

Under the assumption that the two phrases are not represented as flat structures, a syntactic agreement relation between the noun *boys* embedded in the nominal phrase, and the verb *play* embedded in the verb phrase could not be described by strictly local relations, unless the information about the grammatical number were shared by the respective lexical items and their phrasal projections.

The recursive effect of a Head Feature Principle allows such local relations to be extended to the domains within which head features are projected from the lexicon (henceforth the *head domain*).[5] Thus, the requirement of treating certain relations, such as *agreement* or *selection*, as *local* relations could be seen as the ultimate "raison d'être" for head features. Conversely, the declaration of certain properties as head features could be given an independent motivation through an appropriately chosen set of strictly local constraint relations, such as agreement or selection, i.e., the criterion would be that only those features qualify as head features which are crucially involved in these relations.

Still, the question remains whether the theory could be strengthened to the degree that certain features *must* be treated as head features. I.e., could one strengthen the above requirements such that, if a feature is percolated from a lexical anchor to a maximal projection for the purpose of providing a selectional or agreement constraint, it must be a head feature? Of course, we have to accept that selectional

[5]To our knowledge this notion was first introduced by [Engelkamp et al. 1992].

constraints may also involve genuine local features which are manifested only at the level of the selected projection and which have conflicting values at different levels of a head domain.

However, there is nothing to prevent us from exploiting and introducing other mechanisms for the percolation of selectionally relevant features. One example for such an alternative is the [Pollard and Sag 1994] treatment of agreement features, such as `GENDER`, `NUMBER` or `PERSON`, as semantic features transported by virtue of the semantics principle. While this issue touches on the question of the borderline between syntactic and semantic selection, there is the more crucial case of functional categories in *head-marker structures* in [Pollard and Sag 1994]. For these constructions, it is explicitly claimed that a non-head daughter contributes a syntactic feature which is relevant for the selection of the maximal projection. It will be one of the central concerns of our analysis of functional categories presented below to provide an alternative by which this latter instance of selection based on a non-head feature can be eliminated.

To summarize: It can be argued that one central aspect of the relation between heads and their phrasal projections is bound to an issue of locality, in the sense that this relation allows an extension of strictly local domains to *head domains*. These domains are motivated among others by the fact that some properties of a lexical head have to be percolated to phrasal projections to allow the description of certain relations such as agreement or selection as strictly local relations. Selection and agreement, on the other hand, could be taken as necessary (but quite obviously not sufficient) conditions for determining a set of head features. In the further discussion we will show up to what degree these relations could be tightened such that may involve only head features, besides genuinely `LOCAL` properties (which are not projected at all) or non-syntactic properties.

3.3.2 Heads and Grammatical Functions

The second relation crucially involving the notion head, the relation *X-of-head*, concerns the definition of grammatical relations such as *complement-of-head* or *modifier-of-head*, which express some dependency relation between a non-head function and a head. While the notion of heads as designated constituents of phrasal projections appears to be more in the focus of the phrase-structure-oriented tradition of linguistics, this second aspect of the notion head plays a more prominent role in the dependency-oriented tradition.

Two very fundamental aspects have to be kept in mind when talking about the relation between heads and other grammatical functions. Firstly, the grammatical functions involved in such a relation have in common that their presence is *licenced* in a very abstract sense through the presence of a head. This trivially means that for example a subject can only be a subject relative to a head, or that a phrase can only be assigned a modifier function if it modifies some head.[6]

Secondly, the postulation of a set of grammatical functions makes sense only if their relation to the head of a phrase cannot be collapsed with a more general functor-argument or *selection* relation, as we find it, for example, in categorial grammar. If heads are always selectors or functors and non-heads always arguments, which are being selected, there is no need for a separate notion of head. This means that heads have to be primarily motivated, defined and constituted through the above relation head-of-a-phrase involving the projection of head features. The functor-argument or selectional relations may then well cut across head–non-head relations and can be defined differently for different instances of the relation X-of-head. To be more precise, the relation we are talking about here should be actually called X-of-head-of-Y, where X is some grammatical function defined relative to the head and Y some phrase which is projected to the head.

Given these general assumptions, we will discuss in the following how grammatical functions and selectional relations are defined and expressed in HPSG and in other theories. Since selection forms the basis for a number of more specific relations, that may hold between heads and non-heads, such as for example government, agreement or semantic composition, we will also address these aspects in more detail.

Grammatical Functions

Grammatical functions can be basically defined and represented in two different ways. On the one hand, there are theories in which grammatical functions (other than head) are treated as derivative notions, which are defined on pure phrase structure configurations. Another, presumably older tradition, treats grammatical functions as primitives of the descriptive vocabulary. Prominent instances of these two extreme forms are pure $\overline{\mathrm{X}}$-theory and LFG, respectively. HPSG, it can be argued, takes a position somewhere in between.

In theories based on some version of $\overline{\mathrm{X}}$-syntax, grammatical functions are typically secondary notions, which are defined through tree-geometrical relations or phrase structure configurations: a complement is defined as sister to a lexical head X, adjuncts are characterized as sisters to a head projection $\overline{\mathrm{X}}$, specifiers or subjects

[6] This is actually not quite as trivial as it appears, as the example of early LFG shows. In [Bresnan 1982c][42], *functional coherence* is defined such that "every meaningful functional argument of the type specifiable by a lexical form must be bound to a predicate argument", i.e., a subcategorizable function can only appear if there is a head which is subcategorized for it. This excludes the possibility that any of the complement functions appear on their own. However, it does not exclude the existence of structures consisting of minor categories or of adjuncts only, since neither of these are bound to the subcategorization of a head.

are (non-head-) daughters of some maximal projection etc. Since there is no other way to characterize grammatical functions, the respective phrase structure configuration has to be created under all circumstances to distinguish a specific function. For example, a specifier cannot simply be a daughter of XP and sister to a lexical category X, because this would lead to an ambiguity between the specifier function and the complement function. As a consequence, a unary derivation or vacuous projection has to be created, which first raises the lexical category to a level $\overline{\mathrm{X}}$ which only then can be combined with a specifier (or subject) phrase. This assumption, a "Uniform Constituent Hypothesis", not only led to the introduction of such vacuous, unary projections, but also to a multitude of empty terminals (among others for the representation of non-local dependencies), as well as to the postulation of functional categories like INFL or COMP (and their appropriate projections) which do not necessarily have a reflex in the surface string.[7]

By incorporating insights from the Dependency and Relational Grammar tradition, Lexical Functional Grammar explicitly decouples the definition of grammatical functions from phrase structure configurations. At the same time LFG envisages a much more detailed differentiation of grammatical functions, such as subject, direct and indirect object, sentential or verbal complements, predicative and non-predicative adjuncts etc., which are considered as primitives of the theory. This not allows to eliminate certain string-vacuous functional categories and their projections, such as INFL or COMP, from the theory, but also offers a greater freedom in describing classes of languages with a varying degree of configurationality. Thus, for configurational languages with a more or less fixed word order it is suggested to maintain the $\overline{\mathrm{X}}$-convention on the level of phrase structure and to annotate the relevant configurationally defined positions with the corresponding grammatical functions. For non-configurational languages with a higher degree of word order variations, the $\overline{\mathrm{X}}$-convention is practically abandoned and grammatical functions are defined on the basis of functional annotations, which may take into account features such as `CASE` etc.[8]

As mentioned, traditional HPSG takes a position somewhere in between these two approaches. On the one hand it still maintains rudimentary distinctions of the $\overline{\mathrm{X}}$-convention by employing features such as `LEXICAL`, which allow to distinguish between lexical and phrasal projections of a head, or by making reference to the saturation of selectional features. Thus, at least earlier versions of HPSG have in

[7]In early transformational grammar these string-vacuous representations were constrained at the level of surface structure at the price of transformations such as the "Tree Pruning Convention", which eliminated all non-branching elements in a tree, or by the fact that movement transformations did not leave traces. In later versions of generative grammar in the Chomskyan tradition, such transformations were excluded by the requirement that no structure must be eliminated in a derivation. See also [Haider and Netter 1991] on the distinction between derivational and representational realizations of this assumption.

[8]Cf. above all [Bresnan 1982a].

common with configurational approaches that functions such as *subject* were not explicitly marked but determined through a specific configuration.

On the other hand, HPSG chooses to label non-head daughters explicitly as complements (COMP-DTR), adjuncts (ADJ-DTR) etc. This means that many of the grammatical functions distinguished in LFG are subsumed under very abstract and general labels in traditional HPSG, although there is already some tendency to mark some of the more central functions, such as subject or direct object more explicitly.[9] One of the main arguments for choosing such a comparatively small set of abstract grammatical functions, is that this allows to reduce the number of rule schemata to an almost equally small set.[10] In our discussion below (section 3.4), we will show how this set could be simplified and reduced even further.

This combined approach of HPSG bears the potential for more expressivity if it comes to the treatment of word order variations, which in configurational theories have to be reduced to the basic phrase structure configurations. However, some of the restrictions which at first sight appear to come for free in strictly $\overline{\mathrm{X}}$-oriented approaches, have to be compensated for by other mechanisms of the theory. Just as in the $\overline{\mathrm{X}}$-tradition, this has led to the postulation of unary derivations, empty heads, disjunctive listing of values and other devices which allow to maintain a uniform constituent hypothesis at the price of introducing unconstrained mechanisms. As another issue, we mention the problem of defining intermediate projections for as adjunction sites. Many of these devices have in common that they attempt to replay phrase structure derivations on the level of feature structures, instead of breaking down the information contained in a phrase structure tree, such that it can be represented in a more compact and non-redundant form. Thus, in our view there is no need to apply a unary rule which turns a lexical item into a phrase, if the same generalizations can be captured in a less redundant way. It will be one of our objectives below to show how some of these devices, which we view as anachronistic in a constraint-based framework, can be eliminated and replaced by descriptive mechanisms which exploit the simplifying power of feature structures.

Selection

As mentioned, grammatical functions are typically defined relative to heads, with the notion of selection playing a major role. However, theories also differ substantially in as far as not all grammatical functions are involved in selectional relations and not all types of head–non-head relations follow the same selectional patterns.

[9]Cf. amongst others [Borsley 1987], [Borsley 1989] and [Kathol 1994].

[10]Of course, one could argue that if the size of the rule set is a determining criterion, then categorial grammar would clearly win. However, here it comes into play that the grammatical functions and rule schemata in HPSG are associated with an elaborate system of selection defining different relations between heads and other grammatical functions.

Compared to other frameworks, it can be argued that HPSG has one of the most comprehensive and most diversified systems of selection. In principle, all grammatical functions in HPSG are either selected by the `HEAD-DTR` or they select the `HEAD-DTR` themselves. The cases where the head selects the non-head, can be split into those where the non-head is a complement, which is selected through the `SUBCAT` mechanism, and those where the selected function is not an argument of the head, as it is the case for dislocated elements in unbounded dependencies, where the so-called `FILLER-DTR` is selected through a `SLASH` feature on the head. Situations where a non-head selects the head we find in the case of adjuncts (selecting the head through a `MOD` feature), or specifiers and markers (selecting through a `SPEC` feature). In contrast to the latter, the former kind of selection is not limited to a single instance, but may occur arbitrarily often with respect to the same head. Since practically all rule-schemata involve a `HEAD-DTR`, non-head functions cannot occur on their own and are always bound into a selection relation.

In contrast, LFG foresees only one type of selection, namely subcategorization of grammatical functions based on the semantic form or `PRED` feature of the lexical head. The most important difference is that adjuncts as well as minor categories, such as specifiers or complementizers, are neither selected by the head nor do they select the head explicitly in LFG. Thus, restrictions holding between different types of adjuncts and the head, which can be lexically expressed through the `MOD` feature in HPSG, have to be expressed through categorial distinctions on the level of constituent or phrase structure.

In a different style, categorial grammar also envisages only one type of selection encoded as a functor–argument relation. Categorial grammar differs from HPSG most crucially in that no distinction is made between a (semantically defined) functor or selector and a syntactic head. This has important consequences, for example, for the definition of head-modifier or head-adjunct relations. In HPSG the adjunct functions as the selector of the head, while the head continues to determine the properties of the phrasal projection in a direct way. As a consequence the principle determining semantic composition in HPSG has to contain two clauses, one for the case where the semantic "head" collapses with the syntactic head and one where for the case where it doesn't. On the other hand, in categorial grammar, where the focus is more on uniform treatment of semantic composition, the adjunct has to function as the selector, since otherwise the modified constituent would have to be able to select an arbitrary number of adjuncts as arguments. Therefore any projection of properties of the modified category has to be indirectly encoded on the adjunct as the functor. Thus, adjuncts are typically defined such that they take the categorial properties of their argument and map them onto the resulting category, i.e., nominal adjuncts are defined as N/N, etc. Since there is no independent relation between a modified constituent and the resulting phrase, it would be possible, in principle, for an adjunctive functor to change the category of the modified category and to map it onto a different category. Even if one imposed a generalized category X/X on all adjunctive categories, this still means that different types of adjuncts are distinguished by the types of argument they take but do not have an independent categorization of their own. I.e., adjectives and adverbials modifying nouns are both of type N/N, whereas a prepositional phrase modifying a noun or a verb would have to be marked as N/N or V/V respectively.

What most of the feature-based selection mechanisms have in common is that the respective attributes through which selection is expressed are all *projected from the lexicon*.[11] In HPSG these selectional features are in addition constrained to head domains, with the notable exception of `SLASH`. This means that selection either comes from the head or that it comes into force at the maximal projection of a head. For the features `MOD` and `SPEC` this restriction follows trivially from the fact that they are classified as head features which entails that they have to be percolated to the maximal projection to be satisfied. The selection of complements through `SUBCAT` or equivalent features is limited to a head domain by virtue of the Subcategorization or Valence Principle making crucial reference to the `HEAD-DTR` of a phrase.[12] The `SLASH` feature, employed for the treatment of unbounded dependencies, is projected from the lexicon insofar as it is grounded in empty terminals; by the nature of non-local dependencies its distribution domain of course has to go beyond head domains.[13]

It should be pointed out that as an important consequence of this constraint it is basically impossible that a phrasal projection functions as a selector unless either the lexical head of that phrase or, in the case of `SLASH`, some head contained in the phrase licences the selection.[14] At first sight such a constraint appears to be very desirable, its effects being very close to the Projection Principle in G&B which says that all phrasal structures are to be projected from the lexicon. On the other hand, this restriction can be easily circumvented by introducing an empty node functioning as a head and introducing the respective selection feature. A typical example of this method is the treatment of relative clauses

[11] This constraint is very common to other theories as well; however there are also expressed exceptions, such as the selection of subjects through the category VP in earlier theories of the Government & Binding paradigm.

[12] Note, however, that these domains of subcategorization can be extended to higher head domains, if the subcategorization lists are merged or appended, as it is the case with predicate raising. Cf. [Hinrichs and Nakazawa 1994].

[13] There have been also suggestions for an alternative treatment for unbounded dependencies, in which the `SLASH` feature can be actually projected from non-empty lexical heads. See [Netter et al. 1994].

[14] Note that this constraint does not hold in categorial grammar, where type raising allows to turn a category into a functor independent of the properties of its lexical head.

in [Pollard and Sag 1994] where an empty head is introduced, functioning as a 'mediator' between the WH-Phrase, the body of the relative clause and the phrase to which the relative clause adjoins. We will come back to this very crucial issue in our discussions below.

Selection, Government and Agreement

In the previous section we showed that selection is an abstract relation which in HPSG holds between all types of local head–non-head relations. In HPSG (as well as in other frameworks) this notion of selection goes together with the assumption that some part of the structure of a selected constituent is or can be determined by the selector.

In HPSG, selectional relations are typically represented by including the structure of one daughter, normally its `SYNSEM` value, in another daughter's structure as the value of some *selector feature*, be it as some element of the `SUBCAT` list, or as the value of a `MOD`, `SPEC` or `SLASH` feature. This inclusion, expressed by co-indexation of values, is defined on the basis of structural configurations, introducing the grammatical relations, in combination with the relevant principles.

This means that the selector always has full access to all properties of the selected daughter which are contained in the selected substructure. This access to properties of the selected constituent then forms the basis for the representation of more specific selectional constraint relations, such as government, agreement, or the integration of a semantic structure into the semantics of the selecting daughter. Since selector features are discharged structurally, but have to be introduced lexically, the specific constraint relations always can be determined on a lexical basis, even if additional constraints could be specified by the rule schema.

The question arises, however, what status notions such as government or agreement can be given in a constraint-based framework. Should they be seen as intuitive relations which do not necessarily have a specific formal reflex, are they simply two purely formal relations which are not bound to some empirical phenomena, or are there any principled generalizations which allow to decide between one and the other descriptive device?

Intuitively, *government* can be defined as a relation between two constituents X and Y, with X as the governor imposing on the governed constituent Y the constraint to have some property α. *Agreement*, on the other hand, can be seen as a symmetrical relation such that two constituents X and Y agree in some property α, iff the property must have the same value on both X and Y.

Translated into a constraint-based formalism, government corresponds to a configuration where a selector imposes some constraints on a selected constituent which do not occur in the structure of the selector itself. To be more precise, these constraints do not occur outside the selector feature, or at least they are not invariably linked to or coreferent with the value of the selected constituent.

Agreement, on the other hand, means that there has to be an inseparable connection between a property of the selector and a property of the selected constituent which can be represented by coreference of these properties. In theory, agreement could be either expressed lexically through a selector feature or structurally through a coreference relation introduced by a rule. In HPSG, the second option is practically non-existent, since all structural agreement would have to affect some constraint which is already contained in a selector feature anyway, since the selector feature typically encompasses the entire relevant structure of a locally selected sister. The only exception to this rule would be the mediation of structural agreement through empty heads, as it is suggested for example for agreement between the heads of relative clauses and the WH-pronoun.

Even if at first sight these different feature configurations may appear quite unambiguous, it is not always clear where the boundaries between agreement and government are to be drawn. In some cases the same phenomenon can be captured through what looks formally more like a representation of agreement or of government.

For example, the mutual morpho-syntactic dependencies between subjects and finite verbs can be defined as an agreement relation, such that all finite verbs have a feature encoding their inflectional properties which are then identified with the relevant morpho-syntactic properties of the subject.

However, there is also an alternative approach by which this same phenomenon can be represented as a government relation, if the different forms of finite verbs differ with respect to what type of subject they select, but are not marked for any agreement features themselves. In such an approach, which is taken for example in LFG, there is technically no difference between the government of `CASE` values and the selection of `NUMBER` or `PERSON` values of subjects. In a similar and even more radical vein, the analysis proposed in [Pollard and Sag 1994] suggests that subject verb agreement is basically reduced to the selection of specific properties of the semantic index of the subject.

By eliminating the distinction between these different types of relations, the generalization disappears that in many languages we find the morphological forms of finite verbs to co-vary along the parameters of `NUMBER` and `PERSON` with the morphological marking of their subjects, while they are unaffected by the different `CASE` values selected on complements.[15] Of course, whether or not one may consider this as a problem very much depends on what status the morpho-syntactic

[15]We have to admit that we are not quite able to follow the 'natural' conclusion drawn in [Pollard and Sag 1994][83]:

level is given in a framework. In LFG, for example, this analysis presupposes that the lexicon consists of fully inflected forms or that there is a direct link between the morphological component and the selectional properties of the finite verbs. In HPSG, the assumption of a direct link between morphology and semantics becomes feasible through the sign-based approach in which all levels of representation can be described simultaneously.

Whether what is feasible is also appropriate, clearly cannot be given a rough and ready answer. In fact, our main motivation for raising this issue here is that we will encounter and discuss a similar but converse problem below in connection with declension phenomena in German nominal phrases. Whereas one approach to this phenomenon would be based on the assumption of an agreement relation, we will suggest that the relation is described more appropriately as a government relation, precisely on the basis of the assumption of an independent morpho-syntactic levels of abstraction.

Selection and Locality

A final issue, that is not altogether unrelated to the above problem and that we want to address only briefly in this introductory survey, is the question of locality of selectional constraints.

As we mentioned above, one of the prime motivations for the projection of head features within a head domain is that selection can thus be expressed on a strictly local basis. Through selection, both the local properties of a selected phrase can be determined, as well as those of the lexical head of the selected phrase, as far as they are projected as head features.

However, certain selection mechanisms can also be used (or rather misused) to determine properties which do not only affect the local structure of a selected phrase within its head domain, but also structures well outside this domain.

For example, an adjunct selecting a head through its `MOD` feature could impose constraints on the subcategorization of this head by requiring a certain degree

> Unless verbs 'had their hands on' [...] their subjects' indices, they would be unable to assign semantic roles to their subjects. It is natural to expect, therefore, that verb forms might vary in accordance with the indices associated with particular feature specifications associated with these indices, [...]

This conclusion appears to be based on the assumption that the prerequisites for the representation of semantic predicate argument relations provide also some motivation that the morphological inflection forms vary as a function of the semantic context of a constituent, rather than the syntactic context. At least from the point of view of German we do not see any hard evidence for this assumption.

of saturation. However, if the adjunct not only introduced the restriction that the modified phrase must subcategorize say for a subject but also that this subject must have specific properties, this constraint would affect not only the head domain of the modified constituent but also the head domain of one of its complements. An even more extreme situation would be given if we had a recursive embedding of a selector feature such as `MOD`, i.e., if an adjunct selected a constituent which must have a `MOD` feature, through which a constituent must be selected, which must have a `MOD` feature and so on. Even if this latter example may appear as a rather absurd instance of such a recursion, there is a very concrete precedent for the first type in the form of adjective-noun-determiner selection suggested in [Pollard and Sag 1994][88f].

It should be clear that although all of these potential constructions would have non-local effects in the sense of going beyond the head domains, they are still bounded in contrast to genuine un-bounded dependencies which arise only from structural recursion. At least they are bounded as long as the length of the recursive selection path is finite and cannot be specified by a functional uncertainty expression.

The non-locality in the cases described arises from the fact that the `MOD` feature is a head feature which crosses the boundary of a head domain in the adjunct-head selection mechanism. Note by the way that an iteration of a `SUBCAT` feature could not have such a non-local effect. A "recursive" `SUBCAT` selection of a constituent with a non-empty `SUBCAT` value selecting a constituent with a non-empty `SUBCAT` value etc., would loose its effect after the first iteration. Any further constraints on the specific properties of an element of a selected unsaturated `SUBCAT` list are vacuous, since this element would never be realized anyway.

Limiting the effects of selections to local domains has been a declared goal of HPSG. Evidence for this is the restriction of selection to `SYNSEM` or even `LOCAL` values, which, in contrast to the selection of entire *signs* in the earliest versions, excludes constraints on internal phrase-structural properties of a selected constituent.[16]

However, it is very difficult to see how the described cases could be excluded from HPSG, at least not under the circumstances of the currently given feature geometry and treatment of selection. On the one hand, it may not be desirable to be too restrictive. For example, incomplete projections corresponding to a VP can be selected through requiring the selected constituent to have an unsaturated subject slot, thus exploiting this lack of locality restrictions. On the other hand, we are neither aware that any other selectional properties of a constituent (other than the saturation of the subject slot) will play a role in its selection, nor do

[16]See however [Erbach and Krenn 1994] for an approach to idioms which crucially builds on the possibility to impose constraints on the daughters of a selected constituent.

we accept those cases as convincing where embedded (although not necessarily recursive and unbounded) selection is assumed. We will take up this issue in our analysis of determiners below, where we argue that those cases for which such an embedded selection is assumed can be equally well described by mechanisms obeying stronger locality restrictions.

3.3.3 Summary

In this section, we discussed some very general aspects of the notion *head*, which will have some relevance to the analysis outlined below, from the point of view of phrasal projections of heads and in connection with the definition of grammatical functions and selectional relations. As far as the relation *head-of-a-phrase* and the notion of *head features* are concerned, we compared the HPSG treatment with some other prominent constraint-based frameworks and discussed the possibility of finding independently motivated criteria for determining the set of head features. In connection with the relation *x-of-head*, we sketched the relevance of a concept of head for the definition of different grammatical relations, as well as the different types of *selection* that can be involved in head–non-head relations in contrast to functor-argument-relations. We showed that in HPSG selectional relations form the basis of more specific grammatical relations such as *government* or *agreement*. We finally discussed the formal and theoretical possibility of restricting selectional relations to local head domains.

3.4 Rule Schemata and Word Order

Whereas in the previous sections we were more concerned with the grammar model of HPSG and some general very basic concepts and notions, we will go into the details of the theory in this section. The main focus will be on the definition of the different rule schemata interacting with the principles, specifically the treatment of complementation in connection with word order.

Our objective is above all to develop an alternative approach to the one adopted in traditional HPSG. We will try to show how the approach in [Pollard and Sag 1987] and the first part of [Pollard and Sag 1994] can be reformulated in a framework which is basically constrained to binary rule schemata. While at this point we only want to show an equivalence between the specification of complementation schemata as binary and non-binary rule schemata, we will provide some linguistic motivations for our approach in the discussion of German clause structure below (section 5.2).

We will begin with a presentation of the rule schemata as they were proposed in the main part of [Pollard and Sag 1994] and their suggested revisions based on the work of Borsley and others. In the last section we will show how these schemata can be further modified and reduced to binary rules which could form the basis for parameterization of word order in different languages.

3.4.1 Rule Schemata in Pollard and Sag 1994

As already mentioned, phrase structure configurations in HPSG are determined by the interaction of immediate dominance (ID) schemata, linear precedence (LP) principles and other principles, such as the *Head Feature Principle* (HFP) or the *Subcategorization Principle* (SCP). Since we are primarily concerned here with a reformulation of the treatment of complementation, we will leave aside for the moment the schemata postulated for the treatment of head-adjunct relations, for complementizers (*head-marker schema*) and for non-local dependencies (*head-filler schema*).

In [Pollard and Sag 1994] basically three rule schemata are postulated which are to cover the full range of head complement relations in a universal grammar. These schemata are:

Schema 1 (*Head-Subject Schema*) The `SYNSEM|LOC|CAT|SUBCAT` value is $< \;\; >$ and the `DTRS` value is an object of sort *head-comp-struc* whose `HEAD-DTR` value is a phrase and whose `COMP-DTR` value is a list of length one.

Schema 2 (*Head-Complement Schema*) The SYNSEM|LOC|CAT|SUBCAT value is of length one and the DTRS value is an object of sort *head-comp-struc* whose HEAD-DTR value is a word.

Schema 3 (*Head-Subject-Complement Schema*) The SYNSEM|LOC|CAT|SUBCAT value is $< \ >$ and the DTRS value is an object of sort *head-comp-struc* whose HEAD-DTR value is a word.

For these, as well as to all other headed structures the SCP has to be applied as it is formulated in [Pollard and Sag 1994][34,399]:

(61) **Subcategorization Principle** In a headed phrase, the list value of DTRS|HEAD-DTR|SYNSEM|LOC|CAT|SUBCAT is the concatenation of the list value of SYNSEM|LOC|CAT|SUBCAT with the list consisting of the SYNSEM values (in order) of the elements of the list value of DTRS|COMP-DTRS.

Translated into a feature description this principle can be represented as in (62). In such a representation, the part of the SUBCAT list which is consumed by the rule has to be computed on the basis of a functional constraint. This function, which we call **synsem**, has to construct a list of SYNSEM values out of a list of structures of the type *sign*. Note by the way, that for this principle to be generally applicable, all headed structures which do not fall within the class of head-complement relations still have to introduce an empty-list-valued COMP-DTRS attribute.[17]

$$(62)\quad \begin{bmatrix} \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CAT|SUBCAT} \ \boxed{4} \end{bmatrix} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{HEAD-DTR} & \begin{bmatrix} \texttt{SYNSEM|LOC|CAT|SUBCAT} \ \boxed{4} \bullet \text{synsem}\,(\boxed{2}) \end{bmatrix} \\ \texttt{COMP-DTRS} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

The universal schemata and principles are in addition complemented for English by language specific principles determining linear precedence and by parochial additions to the specifications of the daughters. The two basic linear precedence rules determining the order of head complements in English require that lexical heads precede their complements whereas phrasal heads follow them. In addition it is stipulated that the head features of the first two schemata are marked as non-inverted, whereas the head in Schema 3 must be marked as inverted.

Together with the HFP and SCP these specifications will yield the following feature descriptions for rule schemata:

[17] In the representations below we use the symbol $\bullet$ to represent an append operation, concatenating two lists. The symbol $\circ$ is used for the concatenation of two string values.

Schema 1 covers the combination of a phrasal head with its subject. Instances of such phrasal heads are assumed to be simple VPs and nominal heads requiring a determiner or a possessive phrase, which in simple phrase structure terms corresponds to the rules in (63).

$$
\text{(63)}\quad
\begin{array}{llll}
\mathrm{S} & \rightarrow & \mathrm{NP} & \mathrm{VP} \\
\mathrm{NP} & \rightarrow & \mathrm{Det} & \overline{\mathrm{N}} \\
\mathrm{NP} & \rightarrow & \mathrm{NP} & \overline{\mathrm{N}}
\end{array}
$$

The full representation after conjunction with a language specific LP rule for English ("Phrasal heads follow their complements"), the SCP and the HFP is given in (64). Since two of the relevant values required for the SCP are known in this case the **synsem** function can be computed at compile-time for this schema. For the other two schemata described below, it has to be computed at run-time, since in these cases the length of the `COMP-DTRS` list is unknown at compile-time and can be arbitrarily long.

$$
\text{(64)}\quad
\begin{bmatrix}
\texttt{PHON} & \boxed{2} \circ \boxed{1} \\
\texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CAT} & \begin{bmatrix} \texttt{HEAD} & \boxed{3}\begin{bmatrix}\texttt{INV} & -\end{bmatrix} \\ \texttt{LEX} & - \\ \texttt{SUBCAT} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \\
\texttt{DTRS} & \begin{bmatrix}
\texttt{HEAD-DTR} & \begin{bmatrix} \texttt{PHON} & \boxed{1} \\ \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CAT} & \begin{bmatrix} \texttt{HEAD} & \boxed{3} \\ \texttt{LEX} & - \\ \texttt{SUBCAT} & \left\langle \boxed{4} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\texttt{COMP-DTRS} & \left\langle \begin{bmatrix} \texttt{PHON} & \boxed{2} \\ \texttt{SYNSEM} & \boxed{4} \end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Schema 2 is meant to describe all those cases where a lexical head combines with all its non-subject complements and yields as a result a phrase which still requires one complement, the subject. Some possible instances in terms of phrase structure rules are given in (65). Note that all of these schemata also include the case where a lexical category requires only one complement, i.e., where the effect of the schema is equivalent to a derivation with a unary rule.

$$
\text{(65)}\quad
\begin{array}{lll}
\mathrm{VP} & \rightarrow & \mathrm{V\ NP^{*}\ PP^{*}} \\
\mathrm{AP} & \rightarrow & \mathrm{A\ (PP)} \\
\overline{\mathrm{N}} & \rightarrow & \mathrm{N\ (PP)}
\end{array}
$$

In combination with the HFP, the SCP, and the LP principle specific for English (" Lexical heads must be head initial") we obtain the feature description in (66). In this representation the `SYNSEM` values of the `COMP-DTRS` list are concatenated in that order and then appended to a singleton list under the `SUBCAT` value of the `HEAD-DTR`. Different linear orders of the substrings (i.e., values of the `PHON` attribute) corresponding to the individual elements of the `COMP-DTRS` list are supposed to be captured by the function **order**. Similar to **synsem**, this function has to extract the individual `PHON` values, order and concatenate them according to the LP rules determining the order of non-subject complements.

$$
(66)\quad \left[\begin{array}{ll}
\texttt{PHON} & \boxed{1} \circ \text{order}(\boxed{2}) \\
\texttt{SYNSEM} & \left[\begin{array}{ll} \texttt{LOC|CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \boxed{3}\left[\begin{array}{ll}\texttt{INV} & -\end{array}\right] \\ \texttt{LEX} & - \\ \texttt{SUBCAT} & \left\langle \boxed{4} \right\rangle \end{array}\right]\end{array}\right] \\
\texttt{DTRS} & \left[\begin{array}{ll} \texttt{HEAD-DTR} & \left[\begin{array}{ll} \texttt{PHON} & \boxed{1} \\ \texttt{SYNSEM} & \left[\begin{array}{ll}\texttt{LOC|CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \boxed{3} \\ \texttt{LEX} & + \\ \texttt{SUBCAT} & < \boxed{4} > \bullet\, \text{synsem}\,(\boxed{2}) \end{array}\right]\end{array}\right]\end{array}\right] \\ \texttt{COMP-DTRS} & \boxed{2}\end{array}\right]
\end{array}\right]
$$

The third schema was originally intended to capture above all the different cases of auxiliary inversion in English. These constructions are thus accounted for by a "flat" analysis in which the inverted auxiliary (or modal) combines in one go with the subject and a VP complement. The lexical head of this rule, the auxiliary would require the VP complement to have one more element left on its `SUBCAT` list. To control the application of this rule, it is presupposed that all non-auxiliaries are marked as [`HEAD` [`INV` –]], whereas all auxiliaries (and modal verbs) are assumed to be unmarked in this respect.

$$(67)\quad \overline{\text{S}} \rightarrow \text{V} \quad \text{NP} \quad \text{VP}$$

For the ordering of the `PHON` values of the head and complement daughters the same LP principle applies as in the case of Schema 2, namely that lexical heads have to precede their complements.

$$
(68)\quad \left[\begin{array}{ll} \texttt{PHON} & \boxed{1} \circ \text{order}(\boxed{2}) \\ \texttt{SYNSEM} & \left[\begin{array}{ll}\texttt{LOC|CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \boxed{3}\left[\begin{array}{ll}\texttt{INV} & +\end{array}\right] \\ \texttt{LEX} & - \\ \texttt{SUBCAT} & \langle\ \rangle \end{array}\right]\end{array}\right] \\ \texttt{DTRS} & \left[\begin{array}{ll} \texttt{HEAD-DTR} & \left[\begin{array}{ll} \texttt{PHON} & \boxed{1} \\ \texttt{SYNSEM} & \left[\begin{array}{ll}\texttt{LOC|CAT} & \left[\begin{array}{ll}\texttt{HEAD} & \boxed{3} \\ \texttt{LEX} & + \\ \texttt{SUBCAT} & \text{synsem}\ (\boxed{2})\end{array}\right]\end{array}\right]\end{array}\right] \\ \texttt{COMP-DTRS} & \boxed{2}\end{array}\right]\end{array}\right]
$$

Although apparently originally intended for this quite specific construction, this schema turned out to be essential for covering also another major set of phrase structure configurations—the combination of a (lexical) preposition with its complement NP, described by a phrase structure rule such as (69).

(69) PP $\rightarrow$ P NP

Under the assumption that the head of a PP is typically a lexical category, the phrasal projection from a preposition cannot be derived by Schema 2 since this rule requires one element to be left on the `SUBCAT` list of the resulting category. However, it cannot be derived by the Schema 1 either, since this rule requires the head to be phrasal. Even if one derived a phrasal projection $\overline{\text{P}}$ by means of a unary rule application, it would still yield the wrong result since phrasal heads (in English) have to follow their complements.

Thus, the only way to licence the combination of a lexical head which requires only a single complement and does not take an additional subject to its left, appears to be through Schema 3. This may not be such a big surprise at first sight, since this schema appears to be formulated in a sufficiently general way that it should not be too unexpected that prepositional phrases come within its range.

It turns out as an odd feature of this analysis, however, that prepositions have to be marked as [`INV` +] in their `HEAD` features. If this were not the case, there would be nothing to prevent the application of rule Schema 2 to prepositions, which would first yield a unary expansion. This $\overline{\text{P}}$ constituent would then form a phrasal head which according to Schema 1 could be combined with its nominal complement to the left.

Thus, the suspicion arises that what originally looked like a very natural and linguistically fairly well motivated feature marking the inversion of auxiliaries turns more and more into a diacritic feature which can be used to trigger certain rule applications. This suspicion is corroborated further if one takes into account the different possibilities to describe nominal constructions with optional determiners, such as plural or mass nouns. Since in the [Pollard and Sag 1994] approach it is assumed that determiners are subcategorized for by the noun, there are basically two options for handling these cases. The first one would involve the introduction of empty determiners, the second one is based on a disjunctive subcategorization in which the determiner does or does not appear on the SUBCAT list. Assuming that nouns could also be subcategorized for other complements, such as prepositional phrases or clausal constituents, we are confronted with the following situation.

Under the first analysis, involving an obligatory, potentially empty determiner slot, all nouns have to be marked as [INV −]. This is necessary to prevent a derivation, where all complements (determiners and other) appear to the left of the noun by virtue of Schema 3. For those cases, where the noun does not subcategorize for any other complements, a unary derivation to an $\overline{\text{N}}$-constituent by Schema 2 would be enforced which would have to be followed by an application of Schema 1 with the determiner being potentially empty.

However, while such a derivation might still be linguistically appealing to strict followers of an $\overline{\text{X}}$-theory, the second variant involving disjunctive subcategorization appears to require even more twists in the lexical encoding of nouns. Under this analysis, the noun has to be marked as [INV −] if a determiner is subcategorized for, involving again a (potentially unary and vacuous) application of Schema 2 and then Schema 1, as above. For the case where no determiner but only some other complement is required, it is Schema 3 that has to be applied, and the noun form has to be marked as [INV +] accordingly.

3.4.2 Revised Treatments of Valence

The inadequacy of the [Pollard and Sag 1994] approach with respect to prepositional phrases was already pointed out by [Borsley 1987]. In the revisions of their framework in chapter 9, [Pollard and Sag 1994][406ff] therefore suggest to follow Borsley and modify the complementation mechanism such that complements are not all treated indiscriminately by the same subcategorization mechanism. Instead, complements which can be assigned the grammatical function of a subject are separated out and given a different status than other complements. Technically this is expressed by splitting the traditional SUBCAT list into two lists; one list, which is constrained to the length one, represents the subject, another, COMPS, the rest of the complements.

$$(70)\quad \left[\texttt{SYNSEM|LOC|CAT}\begin{bmatrix}\texttt{HEAD} & [\;] \\ \texttt{SUBJ} & \langle[\;]\rangle \\ \texttt{COMPS} & \langle ...\rangle\end{bmatrix}\right]$$

To represent a 'strictly' configurational language like English it then suffices to have two rule schemata, one for cancellation of the subject and one for cancellation of the complements. These two schemata are given the following definition in [Pollard and Sag 1994][408].

Schema 1 (*revised version*) A phrase with DTRS value of sort *head-subj-struc* in which the head daughter is a phrasal sign.

Schema 2 (*revised version*) A phrase with DTRS value of sort *head-comp-struc* in which the head daughter is a lexical sign.

The traditional Subcategorization Principle has to be revised accordingly and is specified as a more general *Valence Principle* which ranges over a set of features (`SUBJ` and `COMPS` in this case) describing the valence of a lexical or phrasal constituent.

(71) **Valence Principle (ValP)** In a headed phrase for each valence feature F, the F value of the head daughter is the concatenation of the phrase's F value with the list of `SYNSEM` values of the `F-DTRS` value.

In terms of feature descriptions, these revised rule schemata together with the Valence Principle can be spelled out as shown in (72) and (73).

(72) *head-subj-schema*

$$\left[\begin{array}{ll} \texttt{SYNSEM} & \left[\begin{array}{ll} \texttt{LOC|CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \boxed{1} \\ \texttt{LEX} & - \\ \texttt{SUBJ} & \langle\,\rangle \\ \texttt{COMPS} & \boxed{3} \end{array}\right] \end{array}\right] \\ \texttt{DTRS} & \left[\begin{array}{ll} \texttt{HEAD-DTR} & \left[\begin{array}{ll} \texttt{SYNSEM|LOC|CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \boxed{1} \\ \texttt{LEX} & - \\ \texttt{SUBJ} & \langle \boxed{2} \rangle \\ \texttt{COMPS} & \boxed{3}\langle\,\rangle \end{array}\right] \end{array}\right] \\ \texttt{SUBJ-DTRS} & \left\langle \left[\begin{array}{ll} \texttt{SYNSEM} & \boxed{2} \end{array}\right] \right\rangle \end{array}\right] \end{array}\right]$$

(73) *head-comp-schema*

$$\left[\begin{array}{ll}
\texttt{SYNSEM} & \left[\begin{array}{ll}\texttt{LOC|CAT} & \left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{LEX} & -\\ \texttt{SUBJ} & \boxed{2}\\ \texttt{COMPS} & \langle\,\rangle\end{array}\right]\end{array}\right]\\
\texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\begin{array}{ll}\texttt{SYNSEM|LOC|CAT} & \left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{LEX} & +\\ \texttt{SUBJ} & \boxed{2}\\ \texttt{COMPS} & \text{synsem}(\boxed{3})\end{array}\right]\end{array}\right]\\ \texttt{COMP-DTRS} & \boxed{3}\end{array}\right]
\end{array}\right]$$

To the representation of Schema 1 we added a restriction that the value of `COMPS` on the `HEAD-DTR` in (72) has to be empty. Under certain assumptions, this constraint can be induced from the fact that any complement on the `COMPS` list must be consumed in a schema with a lexical head, i.e., through Schema 2. Since Schema 2 cannot be applied after Schema 1, but only before Schema 1, it would follow that the input to Schema 1, has to have an empty `COMPS` list. The only case in which this additional assumption about the representation of Schema 1 would not be valid is the unlikely situation that a lexical item is subcategorized for a constituent with an empty `SUBJ` list and a non-empty `COMPS` list.

In Schema 2 on the other hand, it simply follows from the Valence Principle that the phrase's `SUBJ` value is coreferent with its `HEAD-DTR`'s `SUBJ` value. The most important change with respect to the original version is that the value of the `SUBJ` feature is unspecified and that it is not constrained to a list of length one anymore. Whether this value is eventually instantiated to a list of length one or to an empty list fully depends on the lexical head of the schema. While the `SUBJ` value in the case of verbal heads would be a non-empty list, it can also be an empty list in the case of (lexical) prepositional heads.

The problem of determiner-noun-combinations, as we described it above, is solved by a related mechanism. Since for various reasons the assumption that determiners are subjects is rejected [Pollard and Sag 1994][358ff], a third valence feature `SPR` is introduced, which in principle allows to factor out determination from noun complementation.

In this approach, the linear precedence rules for English, as they were given above, remain unaffected: Lexical heads (e.g., prepositions and verbs) precede their non-subject complements, while phrasal heads (e.g., VPs) follow them. Note, that this preservation is closely connected to the fact that intransitive items, i.e., lexical items which only subcategorize for a subject, still have to be submitted to a unary derivation through Schema 2.

However, the question is up to what degree these schemata could be regarded as universal schemata. One of the assumptions which appears to be 'hard-wired' in this configuration of schemata (if its taken to be universal) is that all languages have to reflect the division of subjects and other complements in their word order, i.e., that there is no language in which subjects and other complements occur in an arbitrary order.

To capture such phenomena, the revision of the framework provides the possibility to exploit the effects of a third schema (for aux-inversion) together with a mechanism of complement raising. A second option, which has been suggested by [Borsley 1987] for the treatment of VSO configurations in Welsh, is to postulate a kind of lexical rule which has the effect of removing the subject from the `SUBJ` list and grouping it with the other complements for certain constructions. In languages with even more relaxed word order, the subject would then simply figure as one of the complements on the `COMPS` list and could accordingly appear in any word order position with respect to other objects.

Since we will come back to this issue in great detail in our discussion of finite verb position in German and since here we are mainly concerned with general aspects of the grammar model of HPSG, we raise and briefly sketch only some of the arguments against such a position.

First, as we already pointed out, a relaxation of constraints on the universal schemata to the effect that languages may freely choose among the set of schemata has the effect that one abandons any criterion on what schemata are parochial to a small subset of languages and what are genuinely universal. Even if it could be argued that all schemata may be required for different types of VSO in Welsh, it is still not clear whether this argument carries over to other languages. Since a revised and relaxed application of Schema 2 and/or Schema 3 subsumes the effects of Schema 1, one will either end up with spuriously ambiguous derivations or one will have to exclude Schema 1 from the language particular grammar.

Second, such an approach partly implies that unless the distinction between subjects and other complements is reflected in word order regularities this distinction is irrelevant. However, it may by all means make sense to maintain such a distinction even in free word order languages for other purposes, say control, for the definition of non-finite VPs, for preferences over rule applications or for marking certain word order variations.

Another side effect of the definition of these schemata, in particular the assumption that only lexical heads combine with elements of `COMPS`, is that adjuncts can only adjoin to the left or right of a VP, or alternatively to a completely saturated sentence. Again this is an assumption which may be motivated for a large part of English but does not necessarily hold for other languages.

In the following section we will therefore suggest a different formulation of the two schemata which could account for a wider class of languages and which makes language specific variations the subject of individual parameterizations in the sense of further specification of values.

3.4.3 Universal Schemata and Parameterization

The revision of the definition of rule schemata for complementation as we suggest it below will have the following properties:

- The function of rule schemata is to describe elementary grammatical relations, such as head-subject, head-complement, head-adjunct etc. The vacuous, unary application of a rule schema which describes nothing but the relation *head-of* does not fall under this definition. Unary rule applications (outside the lexicon) are therefore excluded.
- Universal rule schemata are in principle binary branching, i.e., the value of the different DTRS describing grammatical functions (SUBJ-DTR, COMP-DTR, etc.) are of type *sign* rather than *list-of-sign*.[18]
- Linear precedence constraints have nothing (or little) to do with the lexical vs. phrasal status of heads. Instead, they are formulated on the basis of head–non-head relations, taking into account the degree of saturation and the part of speech or category of the head.

Universal Schemata

We begin with a maximally general and hypothetically universal reformulation of Schema 1 and 2, which abandons all reference to the lexical or phrasal status of the `HEAD-DTR` and where all daughters, head or non-head, are specified as type *sign*.

Schema 1 (*universal*) A phrase with DTRS value of sort *head-subj-struc*.

Schema 2 (*universal*) A phrase with DTRS value of sort *head-comp-struc*.

These schemata would be matched by a revised Valence Principle which also ranges over a set of valence features (`SUBJ` and `COMPS`), but takes into account that only one element out of the list value of each valence feature can be consumed within a local configuration.

[18] An exception to the number of branches could be the rule for auxiliary inversion in English which could consist of a ternary branching structure consisting of a `HEAD-DTR`, `SUBJ-DTR` and `COMP-DTR`. Nevertheless, even here the values are not lists but simple signs.

(74) **Valence Principle** (*revised version for binary rules*) In a headed phrase for each valence feature F, the F value of the head daughter is the union of the phrase's F value and the `SYNSEM` value of the value of the `F-DTR`.

As a technical consequence of the assumption that the `F-DTR` value is a simple feature structure of the type *sign*, the relevant `SYNSEM` values (and likewise the `PHON` values) do not have to be computed from the `COMP-DTRS` list anymore, but can be referred to directly.

Under the assumption that there can be only one argument functioning as a subject, Schema 1 would then be spelt out as in (75), where the `COMPS` values of head and mother are coreferent, while the `SUBJ` value of the head daughter is consumed.

(75) *head-subj-schema & ValP*

$$\begin{bmatrix} \mathtt{SYNSEM} & \begin{bmatrix} \mathtt{LOC|CAT} & \begin{bmatrix} \mathtt{HEAD} & \boxed{1} \\ \mathtt{SUBJ} & \langle\ \rangle \\ \mathtt{COMPS} & \boxed{3} \end{bmatrix} \end{bmatrix} \\ \mathtt{DTRS} & \begin{bmatrix} \mathtt{HEAD\text{-}DTR} & \begin{bmatrix} \mathtt{SYNSEM|LOC|CAT} & \begin{bmatrix} \mathtt{HEAD} & \boxed{1} \\ \mathtt{SUBJ} & \langle \boxed{2} \rangle \\ \mathtt{COMPS} & \boxed{3} \end{bmatrix} \end{bmatrix} \\ \mathtt{SUBJ\text{-}DTR} & \begin{bmatrix} \mathtt{SYNSEM} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Conversely in the second schema, the `SUBJ` values of head and mother are co-indexed, while the `SYNSEM` value of the `COMP-DTR` is subtracted from the value of the `COMPS` list, indicated here by a very general function **add** which inserts an item into a list of items.

(76) *head-comp-schema & ValP*

$$\begin{bmatrix} \mathtt{SYNSEM} & \begin{bmatrix} \mathtt{LOC|CAT} & \begin{bmatrix} \mathtt{HEAD} & \boxed{1} \\ \mathtt{SUBJ} & \boxed{2} \\ \mathtt{COMPS} & \boxed{3} \end{bmatrix} \end{bmatrix} \\ \mathtt{DTRS} & \begin{bmatrix} \mathtt{HEAD\text{-}DTR} & \begin{bmatrix} \mathtt{SYNSEM|LOC|CAT} & \begin{bmatrix} \mathtt{HEAD} & \boxed{1} \\ \mathtt{SUBJ} & \boxed{2} \\ \mathtt{COMPS} & \mathrm{add}(\boxed{4}, \boxed{3}) \end{bmatrix} \end{bmatrix} \\ \mathtt{COMP\text{-}DTR} & \begin{bmatrix} \mathtt{SYNSEM} & \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

For lexical items with more than element on the COMPS list, the schema would have to be applied recursively, for lexical items with an empty COMPS list the schema cannot be applied at all.

Language specific parameterization

These two schemata can then be further instantiated according to various language specific parameters. For example, to obtain an [NP VP] structure, as we find it in English, it suffices to add to the first schema the constraint that the COMPS list of the HEAD-DTR must be empty (77). Thus, what could be induced on the basis of the definition of Schema 2 in the original approach, now can be added as a language specific parameter, which leaves it open for other languages not to set the value.

$$(77)\quad \textit{head-subj-schema-E} \;:=\; \textit{head-subj-schema} \wedge \begin{bmatrix}\texttt{DTRS|HEAD-DTR|SYNSEM|LOC|CAT|COMPS}\ \langle\ \rangle\end{bmatrix}$$

The effect of further specifying this value is the same as in the original, since the combination with a subject is allowed only after a head has combined with all its complements. Similarly, the *head-comp-schema* does not have to be further specified with respect to the SUBJ value.

The definition of Schema 2 requires one small addition to the language universal specification, namely that it will be constrained to heads with the feature [INV $-$]. The reason for this will become clear below, when we discuss the derivation of Schema 3.

$$(78)\quad \textit{head-comp-schema-E} \;:=\; \textit{head-comp-schema} \wedge \begin{bmatrix}\texttt{SYNSEM|LOC|CAT|HEAD|INV}\ -\end{bmatrix}$$

Languages, such as German, where a subject may basically appear in any order with respect to other complements, would simply lack a further specification of the COMPS value in Schema 1. Again, the *head-comp-schema* would not make any reference to the possible length or value of the SUBJ, which neatly goes together with the fact that there may be subjectless constructions in German. The fact that non-finite phrases must not contain a subject would be captured most appropriately through the definition of a type *verb-phrase* which has a non-empty SUBJ. Alternatively or in addition, there would be a constraint on the *head-subj-schema* that restricts the head to finite verb forms.

An interesting question is the status of Schema 3 in this approach. So far it has been tacitly assumed that the relation between universal and language specific grammar is a subsumption relation where every language specific schema is subsumed by one universal schema. However, it was never taken into consideration that a language specific rule schema could also be formed by combining two universal schemata. For example, it is rather tempting to treat Schema 3 not as a universal schema, but rather as the unification of Schema 1 and 2. Under the original approach such an assumption would have been impossible, since there the two schemata involve conflicting values for the LEX feature. Since we eliminated all reference to this feature from the two schemata, there is a priori no reason why such a solution should be excluded.

However, whether we can make such an assumption will crucially depend on the precise definition of the individual universal schemata in terms of appropriateness specifications as well as the exact interpretation of the Valence Principle. This principle is intended to have two different kinds of effects. For those valence features for which a corresponding daughter is derived, the value of this feature is to be reduced, for all other valence features the value has to be passed on unaffected to the mother node.

In particular, it has to be clear how the values of F-daughters other than the specific grammatical function which is being reduced in a head complement relation are specified. For example, is the universal *head-subj-schema* specified such that all non-head daughters, other than the one under the feature path DTRS|SUBJ-DTR|SYNSEM, are defined as inappropriate, are all these daughters assigned a specific value, say **nil**, are these feature paths blocked by feature negation, or is it sufficient to define the SUBJ-DTR path as appropriate while leaving the other attributes completely unspecified.

In the latter case (and presumably not only in this case), it might be more appropriate to formulate the Valence Principle along the following lines:

Valence Principle In a headed phrase for each valence feature F, the F value of the head daughter is the union of the phrase's F value and the SYNSEM value of the value of the F-DTR, if this SYNSEM feature is defined as appropriate. Otherwise the value of of a phrase's valence feature is identical to the HEAD-DTRs valence feature.

If such a definition of the Valence Principle were compatible with general assumptions about the application of principles, Schema 3 could easily be interpreted as the unification or intersection of Schema 1 and 2. From the first it would inherit the information that the SUBJ-DTR|SYNSEM path is defined, from the second that the path COMP-DTR|SYNSEM is appropriate. The derived Schema 3 in English would then be simply defined as in (79), and would have the form (80), when expanded together with the Valence Principle, either added to the individual schemata in the Universal Grammar or in the language particular grammar. The addition that we have to make here is that the HEAD-DTR is specified as [INV +], which will allow to constrain the application of this rule to auxiliaries and modals, as well as prevent that the head first combines with some other complement before it consumes the subject and the last complement.[19]

$$\text{(79)}\quad \textit{head-subj-comp-schema-E} \;:=\; \textit{head-subj-schema} \wedge \textit{head-comp-schema} \wedge \begin{bmatrix}\texttt{SYNSEM|LOC|CAT|HEAD|INV} + \end{bmatrix}$$

[19]Ivan Sag (p.c.) drew our attention to a class of constructions which are not easily covered either by this or the more traditional approaches. These constructions arise from verbs which are ambiguous between auxiliary and main verb, i.e., the copula *be* and the main verb *have*, for which inverted structures exist which do not require the support of the auxiliary *do*.

(i) Is Kim here?

(ii) Has Kim more than one cookie?

While these simple constructions are clearly covered by the ternary rule schema, they differ from those with genuine auxiliary, as in the latter case there is an obvious site for adjunction to the main verb.

(iii) [Has Kim [already [been here]]]?

(iv) [Did Kim [already [have more than one cookie]]]?

In the analogous constructions with the main verb inverted such an obvious adjunction site is missing.

(v) Was Kim *in his past life* a teacher?

(vi) Had he *not* several books, when you last saw him?

Analyzing these constructions such that the adverbial is adjoined to the predicative or object (vii) is quite likely to lead to inconsistencies and inadequacies on a semantic level. Choosing a left-branching binary structure will be inconsistent with the assumption that in the *head-subj-schema* the `COMPS` list in English has to be empty.

(vii) [Was Kim [in his past life a teacher]]?

(viii) [[[Was Kim] in his past life] a teacher]?

An alternative analysis, which could be taken into consideration, would therefore build on a binary right branching structure simulating some kind of "head-movement", along the lines we are going to suggest below for German.

(ix) [Was [Kim [in his past life [*eps-V* a teacher]]]]?

(80) *head-subj-comp-schema & ValP*

$$\begin{bmatrix} \texttt{SYNSEM} & \begin{bmatrix}\texttt{LOC|CAT} & \begin{bmatrix} \texttt{HEAD} & \boxed{1}\begin{bmatrix}\texttt{INV} & +\end{bmatrix} \\ \texttt{SUBJ} & \langle\,\rangle \\ \texttt{COMPS} & \boxed{3}\end{bmatrix}\end{bmatrix} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{HEAD-DTR} & \begin{bmatrix}\texttt{SYNSEM|LOC|CAT} & \begin{bmatrix}\texttt{HEAD} & \boxed{1} \\ \texttt{SUBJ} & \boxed{2} \\ \texttt{COMPS} & \text{add}(\boxed{4},\boxed{3})\end{bmatrix}\end{bmatrix} \\ \texttt{SUBJ-DTR} & \begin{bmatrix}\texttt{SYNSEM} & \boxed{2}\end{bmatrix} \\ \texttt{COMP-DTR} & \begin{bmatrix}\texttt{SYNSEM} & \boxed{4}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Finally, what appears to be missing for English in this approach is the restriction that adjuncts typically must not be mixed with elements on the COMPS list but should typically attach to saturated VPs or sentences only (leaving aside cases of heavy NP-shift and some related constructions). This restriction cannot be expressed through Schema 2 anymore, since the original constraint on the HEAD-DTR to be a lexical category is now eliminated. As a consequence a head could theoretically combine with another complement even after it *head-adjunct-schema* has been applied.

However, the expression of this constraint can easily go where it could have been in the first place. It could be located either in the structural schema introducing adjuncts or, preferably, in the MOD specifications of the individual heads of adjuncts. By restricting the HEAD-DTR in a *head-adjunct-schema* to structures with a saturated COMPS list, we will have the same effect as it could be obtained from the original feature [LEX +]. Locating the constraint in the MOD feature would even give us the flexibility to allow adjunction to unsaturated lexical items if necessary.[20] Again, languages such as German would lack this constraint, since adjuncts can basically appear in any order in between complements and subjects.

Parameterization of LP constraints

In the parametrizations above, word order is taken into account only as far as it is reflected in constituency, i.e., in certain configurations the subject is forced

[20] A potentially relevant case could be adjunction of the uncontracted negation particle to inverted auxiliaries as in

(i) Does not everybody know that ...

to appear in a position which is peripheral to the objects. LP constraints, which went beyond this general constraint, were formulated for English in the traditional approaches presented above on the basis of the distinction between lexical and non-lexical heads. In our revised version, this distinction is neutralized —in a *head-comp-schema* but also in a *head-subj-schema* the head can be a lexical or a non-lexical constituent.

We therefore have to find an equivalent formulation. Our preferred option here is to base the constraints on the saturation of heads, specifically the opposition between empty and non-empty `COMPS` lists.

For English, for example, an equivalent formulation would specify that heads with an empty or saturated `COMPS` list follow their arguments (be it subjects and possibly also specifiers). Heads with a non-empty `COMPS` list precede all of their arguments. To put it differently, heads with empty `COMPS` lists are constrained to a head-final position, while heads with non-empty `COMPS` lists are head-initial. This latter case would cover verbs or verbal projections which still look for some object, but also prepositions with a nominal complement. Note, that this formulation also covers the constructions with auxiliary inversion in English, for which we assumed a combination of Schema 1 and 2, resulting in a ternary structure. Since the head in this rule schema at the same time combines with a subject and a complement, it would have a non-empty `COMPS` list and accordingly precedes all of its arguments.

One of the most crucial differences between this formulation and the one given in the [Pollard and Sag 1994] approach is that our approach dispenses with unary rule expansions, which are necessary in the traditional versions described above. These unary derivations were originally motivated almost exclusively by the generalization that heads which follow their subjects had to be phrasal. Thus, if a lexical head (e.g., an intransitive verb) did not require any other complements, it had to be 'raised' to the status of a phrasal constituent (equivalent to a VP) through an application of the *head-comp-schema* with an empty list as the value for `COMP-DTRS`. Since the LP constraints are now formulated on the basis of complementation rather than the lexical–phrasal distinction, such unary rules are now unnecessary. In fact, none of our revised schemata would licence such unary expansions, since we specify the values of the various types of daughters as feature structures of the type *sign* rather than a potentially empty *list-of-signs*.

The LP constraints as we gave them allow us to define a VP with an internal head-initial structure, i.e., SVO. Languages with a head-final VP structure, i.e., an SOV structure, would follow the constructional constraints in the rule schemata, but have an LP constraint which simply requires all (verbal) heads to be head-final. As mentioned, languages such as German would lack the constructional constraints in the schemata. The head position here could be best described by LP constraints which require all verbal and adjectival heads as head-final and all nominal and prepositional to be head-initial.[21]

Other Word Order Constraints

So far we only considered word order constraints which are either entailed by the configurational structure in different languages, i.e., the question whether or not there is a closed VP containing all objects, or what we consider to be hard LP constraints, i.e., the order of heads relative to complements or modifiers. However, there are also certain soft LP constraints which concern the order of objects or adjuncts among each other. For the expression of these constraints different options can be envisaged.

One way to determine the order of elements selected through the `COMPS` list would be to assume different possibilities of reducing the list. If the order of objects can be fully predicted already in the lexicon, it would suffice to indicate this order in the `COMPS` list and then reduce this list by means of a simple **cons** function, which always takes away the first element. For more complex or semi-free orders the reduction of the list would have to be determined by a **union** function, as it was assumed above. Such a function could either be defined such that it takes into account more complex constraints straightaway. For example, one could assume that the order on the `COMPS` list reflects the unmarked order of complements and define the function such that it reduces the `COMPS` list by preferably taking off elements as close to the beginning of the list as possible. Alternatively, the reduction function could be complemented by non-local LP rules, i.e., rules determining the order of elements within a domain which goes beyond trees of depth one. A proposal along these lines is to be found in [Engelkamp et al. 1992].

We are well aware that all of these options would have to be worked out in considerably more detail. However, since these issues are not the central concern of this thesis and since so far there are no fully elaborated proposals for this area in the more traditional HPSG approaches, we will not discuss the problem any further.

3.4.4 Summary

In this section we discussed the traditional approaches to complementation and the potential relationship between universal and language particular rule schemata. We suggested an alternative approach in which basically two universal schemata are postulated, whose function is to introduce local configurations defining grammatical relations such as head-subject and head-complement. We showed how

[21] Such a distinction could draw on the characterization [`V` +] and [`V` −] in the Chomskyan classification of open classes. We will come back to this issue in more detail below in section 4.3 and in section 5.3.

these universal schemata can be parametrized by language specific information thus accounting for the distinction between configurational and non-configurational languages. The language specific schema for auxiliary inversion in English we treated as a derivative schema which is composed of the two universal complementation schemata.

We defined equivalent formulations for LP constraints, which build on the degree of saturation rather than the distinction between lexical and phrasal constituents. As an important consequence, the approach allows to dispense with the stipulation of unary rule derivation, which were apparently motivated almost exclusively by LP constraints building on this distinction.

3.5 Principles and Rule Schemata

In the chapter on the underlying formalism we discussed some general approaches to the representation and interaction of principles and rule schemata. In this section we will give a more precise definition of principles and their mutual dependencies by sorting them into a type hierarchy.

We adopt the revised definition of principles and rules in [Pollard and Sag 1994], according to which also the various rule schemata are organized as partitions of an Immediate Dominance Principle. For our purposed it suffices to restrict ourselves to the three main principles referred to above: the Head Feature Principle (HFP), the Valence Principle (ValP) and the Semantics Principle (SP), as well as the Immediate Dominance Principle (IDP), all of which are repeated in prose in (81). Since we are not concerned with the details of semantic construction we use the simplified version of the SP as it is defined in [Pollard and Sag 1994][53], ignoring the effects of quantifier retrieval.

(81) **HFP** In a headed phrase the values of [`SYNSEM|LOC|CAT|HEAD`] and [`DTRS|HEAD-DTR|SYNSEM|LOC|CAT|HEAD`] are token-identical.

ValP In a headed phrase for each valence feature F, the F value of the head daughter is the union of the phrase's F value and the `SYNSEM` value of the `F-DTRS` value.

SP In a headed phrase, the `CONTENT` value is token-identical to that of the adjunct daughter if the `DTRS` value is of sort *head-adj-struc*, and with that of the head daughter otherwise.

IDP Every headed phrase must obey exactly one of the ID schema.

Our general approach here is to view principles as generalizations over sets of objects in a type hierarchy, rather than as independent implicational constraints which have to be applied to linguistic objects. For example all of the above principles have in common that they are defined only for phrasal signs of the type *headed-phrase*. Thus, we would have basically two choices to represent these principles relative to a type hierarchy. We could either define them in the most general way completely independent of the definition of phrasal signs. Else, we can restrict their applicability right from the beginning by representing them as subtypes of objects of the type *headed-phrase*. We pursue the latter option here and assume that the principles are 'sorted' into the type hierarchy as subtypes to the most specific type which would subsume their antecedent in an implicational representation of the principle.

We begin with the definition of the type *phrasal-sign*, which we assume to be partitioned into *headed-phrase* and *headless-phrase*, both of which define what types of values the DTRS attribute may have.

(82)
$$phrasal\text{-}sign \;:=\; sign \;\&\; \begin{bmatrix} \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CAT} & phrasal\text{-}cat\begin{bmatrix}\;\end{bmatrix} \end{bmatrix} \\ \texttt{DTRS} & dtrs\begin{bmatrix}\;\end{bmatrix} \end{bmatrix}$$

$$headed\text{-}phrase \;:=\; phrasal\text{-}sign \;\&\; \begin{bmatrix} \texttt{DTRS} & dtrs\begin{bmatrix} \texttt{HEAD-DTR} & sign\begin{bmatrix}\;\end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$headless\text{-}phrase \;:=\; phrasal\text{-}sign \;\&\; \begin{bmatrix} \texttt{DTRS} & dtrs\begin{bmatrix} \texttt{HEAD-DTR} & *undef* \end{bmatrix} \end{bmatrix}$$

$$\text{part } (phrasal\text{-}sign) \;\equiv\; headed\text{-}phrase,\; headless\text{-}phrase$$

Since all headed phrases have to obey the HFP we can assume this principle as the only subset of the type *headed-phrase* (83).

(83)
$$HFP := \;\; headed\text{-}phrase \;\& \;\; \begin{bmatrix} \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CAT|HEAD} \; \boxed{1} \end{bmatrix} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{HEAD-DTR|SYNSEM|LOC|CAT|HEAD} \; \boxed{1} \end{bmatrix} \end{bmatrix}$$

$$\text{part } (headed\text{-}phrase) \;\equiv\; HFP$$

The partial lattice rooted in this type and illustrating the inheritance relations of the type definitions developed below is given in figure 3.1.

The class of objects satisfying the HFP also has to obey the three other principles, the SP, ValP and IDP, which for the moment we simply subordinate to the HFP without assigning further constraints to the definitions. We also assume that the type *HFP* is cross-classified by these three principles, i.e., that every object in the set denoted by *HFP* has to be a member of each of its immediate subclasses.

(84) $SP := HFP$

$ValP := HFP$

$IDP := HFP$

$\text{cross } (HFP) \;\equiv\; SP,\; ValP,\; IDP$

If we now take a closer look at these three principles, it appears that the SP and ValP can be broken down into a more specific case and an 'elsewhere' or default case. To begin with the Semantics Principle, it defines for some specific

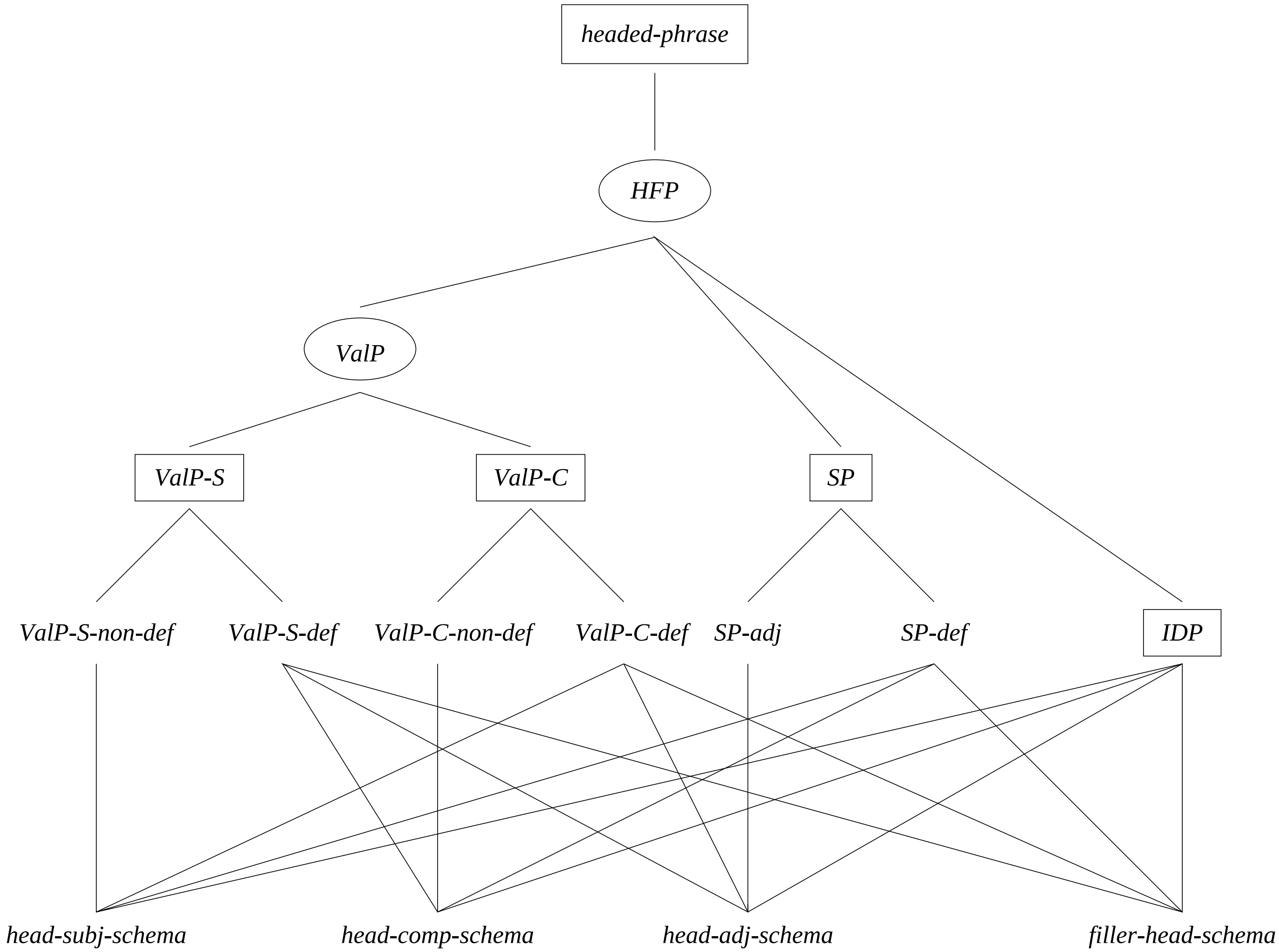


Figure 3.1: Inheritance Hierarchy for Principles and Schemata

structures, namely those for which an ADJ-DTR is defined, that the CONT value of the mother is identified with the adjunct daughter's CONT value, elsewhere the mother's CONT value comes from the HEAD-DTR. To capture this generalization we partition the SP into two subtypes, one which holds for headed phrases with an adjunct (*SP-adj*), and one which applies elsewhere as a default (*SP-def*).

(85) part (*SP*) $\equiv$ *SP-adj*, *SP-def*

$$\textit{SP-adj} := \quad \textit{SP} \; \& \; \begin{bmatrix} \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CONT} \; \boxed{1} \end{bmatrix} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{ADJ-DTR|SYNSEM|LOC|CONT} \; \boxed{1} \end{bmatrix} \end{bmatrix}$$

$$\textit{SP-def} := \quad \textit{SP} \; \& \; \begin{bmatrix} \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CONT} \; \boxed{1} \end{bmatrix} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{HEAD-DTR} & \begin{bmatrix} \texttt{SYNSEM|LOC|CONT} \; \boxed{1} \end{bmatrix} \\ \texttt{ADJ-DTR} & *undef* \end{bmatrix} \end{bmatrix}$$

At first sight, the ValP, as it is defined above, appears to apply uniformly to all headed phrases. Yet, as we already pointed out in the previous section, it is not quite clear how those cases are to be treated where there is no defined SYNSEM value for an F-DTR. Also, the prose of this principle hides a quantification over the variable F which at the same time is meant to refer to part of the attribute name F-DTR and to the respective valence feature. To make this hidden information more explicit, we first of all spell out the Valence Principle, such that all different cases become more perspicuous and such that all features and operations referred to can be safely assumed to be defined.

(86) **Valence Principle** (*elaborate version*)

In a headed phrase of the type *head-subj-structure* the SUBJ-value of the HEAD-DTR is the sum of the phrases SUBJ value and the SYNSEM value of the SUBJ-DTR of the phrase.

In a headed phrase of the type *head-comp-structure* the COMPS-value of the HEAD-DTR is the sum of the phrases COMPS value and the SYNSEM value of the COMP-DTR of the phrase.

Else the F-values of the phrase and the F-values of the HEAD-DTR are token-identical.

To translate this specification into a feature description we then assume that the ValP is cross-classified by two subsets (*ValP-S* and *ValP-C* in (87)), dealing with the two types of complementation. Each of these two subtypes is represented as a partition (88) and (89), which covers on the one hand the specific case, where the valence feature is reduced, and on the other hand the elsewhere or default case, which simply passes on the value of the valence feature to the mother.

(87) cross (*ValP*) $\equiv$ *ValP-S*, *ValP-C*

(88) part (*ValP-S*) $\equiv$ *ValP-S-non-def*, *ValP-S-def*

$$\textit{ValP-S-non-def} := \textit{ValP-S}\ \& \left[\begin{array}{ll}\texttt{SYNSEM} & \left[\texttt{LOC|CAT|SUBJ}\ \langle\ \rangle\right] \\ \texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\texttt{SYNSEM|LOC|CAT|SUBJ}\ \left\langle \boxed{1} \right\rangle\right] \\ \texttt{SUBJ-DTR} & \left[\texttt{SYNSEM}\ \boxed{1}\right] \\ \texttt{ADJ-DTR} & *undef* \\ \texttt{FILLER-DTR} & *undef* \end{array}\right]\end{array}\right]$$

$$\textit{ValP-S-def} := \textit{ValP-S}\ \& \left[\begin{array}{ll}\texttt{SYNSEM} & \left[\texttt{LOC|CAT|SUBJ}\ \boxed{1}\right] \\ \texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\texttt{SYNSEM|LOC|CAT|SUBJ}\ \boxed{1}\right] \\ \texttt{SUBJ-DTR} & *undef* \end{array}\right]\end{array}\right]$$

(89) part (*ValP-C*) $\equiv$ *ValP-C-non-def*, *ValP-C-def*

$$\textit{ValP-C-non-def} := \textit{ValP-C}\ \& \left[\begin{array}{ll}\texttt{SYNSEM} & \left[\texttt{LOC|CAT|COMPS}\ \boxed{3}\right] \\ \texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\texttt{SYNSEM|LOC|CAT|COMPS}\ \boxed{1}\right] \\ \texttt{COMP-DTR} & \left[\texttt{SYNSEM}\ \boxed{2}\right] \\ \texttt{ADJ-DTR} & *undef* \\ \texttt{FILLER-DTR} & *undef* \end{array}\right]\end{array}\right]$$

$$\&\ \text{COND: } \boxed{1} = \text{add}\ (\ \boxed{2}\ ,\ \boxed{3}\ )$$

$$
\textit{ValP-C-def} := \textit{ValP-C} \;\& \left[\begin{array}{ll} \texttt{SYNSEM} & \left[\texttt{LOC|CAT|COMPS}\ \boxed{1}\right] \\ \texttt{DTRS} & \left[\begin{array}{ll} \texttt{HEAD-DTR} & \left[\texttt{SYNSEM|LOC|CAT|COMPS}\ \boxed{1}\right] \\ \texttt{COMP-DTR} & *\textit{undef}* \end{array}\right] \end{array}\right]
$$

In the specification of these types we excluded the possibility that a complement is consumed in a local configuration which also contains an ADJ-DTR or a FILLER-DTR. The cross-classification and partition of these types allows for four different combinations. It can lead to two binary rules in with either a COMP-DTR or SUBJ-DTR is being consumed and the complementary value is being passed on (*ValP-S-non-def* and *ValP-C-def*, and *ValP-S-def* and *ValP-C-non-def* respectively). In addition, we allow the option of a ternary rule in which both types of complements are being consumed, which would correspond to Schema 3 in the traditional approach (*ValP-S-non-def* and *ValP-C-non-def*). The combination of the two defaults *ValP-S-def* and *ValP-C-def* covers the case of all rules which do not involve complementation. Only the latter combination is compatible with the specific case of the Semantics Principle, while the former three can only be combined with the default case. The combination of all default cases of these three principles and sub-principles would be compatible with schemata such as the *head-filler-schema*, involving neither adjunction nor complementation.

We finally have to combine these types with the Immediate Dominance Principle, which already by definition requires a partitioning of the set of objects to which it applies. At this point, we will assume only four abstract schemata here, which illustrate the different types of selection on the basis of the SUBJ, COMPS, MOD and SLASH attributes.

(90) part (*IDP*) $\equiv$ *head-subj-schema*,*head-comp-schema*, *head-adj-schema*, *head-filler-schema*

If we wanted to follow a strict course, we could represent these schemata as subtypes of one exclusive supertype, the *IDP*. In this case the cross-classification of the different principles would have to be computed along the lines we sketched above. However, we can also define these schemata as the appropriate intersection of the various principles yielding the cross-classification 'by definition'.

Thus, the *head-subject-* and *head-complement-schema* can be defined simply as the intersection of the respective non-default and default valence principles, the default Semantics Principle, and the IDP.

(91) *head-subj-schema* := *ValP-S-non-def* & *ValP-C-def* & *SP-def* & *IDP*

(92) *head-comp-schema* := *ValP-C-non-def* & *ValP-S-def* & *SP-def* & *IDP*

Similarly, the *head-adjunct-schema* is composed of the default Valence Principles and the specific part of the Semantics Principle. In addition a constraint is introduced which coindexes the HEAD-DTR's SYNSEM value with the ADJ-DTR's selector feature MOD.

(93)

$$\textit{head-adj-schema} := \textit{SP-adj} \;\&\; \textit{ValP-S-def} \;\&\; \textit{ValP-C-def} \;\&\; \textit{IDP} \;\&\; \left[\begin{array}{ll}\texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\texttt{SYNSEM}\ \boxed{1}\right] \\ \texttt{ADJ-DTR} & \left[\texttt{SYNSEM|LOC|CAT|HEAD|MOD}\ \boxed{1}\right]\end{array}\right]\end{array}\right]$$

We finally specify the *head-filler-schema* as an example which involves the intersection of all default branches of the relevant principles, with the effects of the Non-Local Principle added in a simplified form as a constraint.

(94)

$$\textit{head-filler-schema} := \textit{ValP-S-def} \;\&\; \textit{ValP-C-def} \;\&\; \textit{SP-def} \;\&\; \textit{IDP} \;\&\; \left[\begin{array}{ll}\texttt{SYNSEM} & \left[\texttt{NON-LOC|SLASH}\ \boxed{3}\right] \\ \texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\texttt{SYNSEM|NON-LOC|SLASH}\ \boxed{1}\right] \\ \texttt{FILLER-DTR} & \left[\texttt{SYNSEM}\ \boxed{2}\right]\end{array}\right]\end{array}\right]$$

$$\&\ \text{COND: } \boxed{1} = \text{add} \left(\boxed{2}\ , \boxed{3} \right)$$

To illustrate the structures resulting from these definitions we give the fully expanded form of the *head-subj-schema* in (95).

(95)

$$\left[\begin{array}{ll}\texttt{SYNSEM} & \left[\texttt{LOC|CAT}\ \left[\begin{array}{ll}\texttt{HEAD} & \boxed{1} \\ \texttt{SUBJ} & \langle\ \rangle \\ \texttt{COMPS} & \boxed{3}\end{array}\right]\right] \\ \texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\texttt{SYNSEM|LOC|CAT}\ \left[\begin{array}{ll}\texttt{HEAD} & \boxed{1} \\ \texttt{SUBJ} & \left\langle \boxed{2} \right\rangle \\ \texttt{COMPS} & \boxed{3}\end{array}\right]\right] \\ \texttt{SUBJ-DTR} & \left[\texttt{SYNSEM}\ \boxed{2}\right] \\ \texttt{COMP-DTR} & *\textit{undef}* \\ \texttt{ADJ-DTR} & *\textit{undef}* \\ \texttt{FILLER-DTR} & *\textit{undef}*\end{array}\right]\end{array}\right]$$

3.6 Summary

In this chapter we discussed some of the basic assumptions of HPSG and made some suggestions on how to modify the setup of the theory to account for language specific variations while maintaining the basis of a hypothetical universal grammar. We challenged the traditional definition of universal grammar in HPSG as not providing a sufficient criterion for determining the content of a universal grammar. Instead we argued for a stratified approach in which universality of a concept is determined bottom up in terms of generalizations across different (classes of) languages.

We discussed the notion of *head*, as one of the most central concepts of HPSG and in comparison with other frameworks, in terms of two relations; the relation heads and their phrasal projections which is crucially bound to a notion of *head features* and which allows to define a locality notion of head-domains, and the relation between heads and their sister constituents which serves to define grammatical relations relative to heads. Secondary to this latter relation are notions such as selection, government and agreement which determine specific relations holding between grammatical functions. Quite a number of the issues we touched upon in this section will be taken up below in our analysis of functional categories as specific types of heads.

In section 3.4 we reviewed the HPSG approaches to phrase structure configurations and word order. We pointed out certain inadequacies of these traditional assumptions and set out to redefine the rule schemata for complementation. Our objective here was to generalize the rule schemata, so that they are more likely to qualify for a concept for universal grammar, as well as to derive and parametrize more specific instances of rule schemata for the definition of language specific grammars. We finally showed how most of principles of HPSG can be integrated into a type hierarchy, representing the ontological objects of a grammar, by spelling them out into a format of default and non-default sub-principles.

Chapter 4

Nominal Phrases

4.1 Overview

In this chapter we develop a notion of functional heads in the framework of HPSG which differs from the assumptions about specifiers as they are outlined in [Pollard and Sag 1987] and [Pollard and Sag 1994]. We base our approach on a description of a fragment of German nominal phrases, concentrating on the categorial specifications of determiners, nouns and adjectives, as well as morpho-syntactic relations holding between these categories within a nominal construction.

To set the background for our approach, we discuss in the first section 4.2 what implications the notion of 'head', as we outlined it above in section 3.3, has with respect to nominal phrases. The purpose of this discussion is mainly to clarify what criteria we have, from a theory internal point of view, for deciding on what category—the noun or the determiner—should function as the head of a nominal phrase. The reflex of this question in standard $\overline{\mathrm{X}}$-theory is the distinction between analyses in which the maximal category of a nominal phrase is an NP, a projection of the category N with the determiner functioning as specifier (cf. [Jackendoff 1977]), and those which treat nominal phrases as a DP, a projection of D, where the noun and its adjuncts function as a complement of the determiner (cf. [Abney 1987]). We will address some of the implications and difficulties arising from these approaches if pursued in their purest form.

As an alternative, we present an analysis in sections 4.3 and 4.4 which at first sight may look like a 'DP-analysis'. However, the approach crucially differs from a pure DP- or NP-analysis insofar as it does not require that a nominal phrase be *always* a projection of a category D (or N), since this would commit us to the postulation of empty determiners (or unary rules) in the case of determinerless constructions.

One of the central ideas of our analysis is that categories are defined through two different sets of `HEAD` features which are declared as `MAJOR` and `MINOR` properties, recalling the distinction between major, substantive or lexical categories—such as nouns, verbs, adjectives and prepositions—and minor or functional categories—such as determiners, numerals or complementizers. In contrast to the traditional usage of these notions, we assume that all categories may be defined for substantive and functional properties simultaneously.

To define maximal projections, we introduce a primitive notion of *Functional Completeness* denoting that a category must not or need not combine with a specifier or functional head anymore. This notion, which is encoded as a `MINOR` property, has a status which is similar to the saturation of subcategorization requirements of a head. A maximal nominal projection, for example, is essentially defined through three different properties: full saturation of subcategorization requirements, certain `MAJOR` features which will characterize the category as nominal, and certain

MINOR features which will specify it as *functionally complete*.

Functional categories we then define so that they function as heads of maximal projections subcategorizing for a functionally incomplete category. Thus, a determiner may function as the head of the nominal phrase and is subcategorized for the noun or some nominal projection which has not combined with a determiner. Functional heads operate in a parasitic way on their complements insofar as they identify the head features of their complement with their own, while also introducing the minor properties marking a phrasal projection as functionally complete.

This will allow us to mark certain noun forms, in particular singular count nouns, as functionally incomplete, so that they have to combine with a determiner in order to make up a maximal nominal projection. Mass nouns and plural nouns will be underspecified in this respect, so that they can either function as the complements of determiners or directly as the heads of nominal phrases. As a result, we can derive determinerless constructions without having to make recourse to mechanisms such as empty determiners or unary expansions of the root node of a nominal phrase.[1]

Up to this point, we will have been concerned mainly with the problem of how to define maximal nominal projections and how to relate them to their respective heads. Section 4.5 will discuss the problem of intermediate projections corresponding to the category $\overline{\mathrm{N}}$ in a traditional NP-analysis. The question we address there is as simple as how to define the category to which attributive adjectives as prenominal modifiers must attach. We will show that neither the saturation of subcategorization requirements nor the notion of functional completeness is suited to define such an intermediate projection. To solve the problem we opt for a lexical solution introducing a further MINOR distinction in the lexical categories of determiners and adjectives, rather than a structural solution involving empty determiners or unary rules. Since this problem and its solution could be seen as either a weakness or a strength of our approach, we take up the question again in the very last section where we compare our approach with the analysis suggested by [Pollard and Sag 1994].

While we argued decisively against the postulation of empty terminals functioning as determiners, we take a rather different point of view with respect to empty

[1] Although we developed it quite independently, taking as a starting point the treatment of determiners in [Pollard and Sag 1994], this definition of functional categories turned out to be not quite as novel as we originally thought. A quite similar proposal into which this HPSG version can be translated almost directly was made by Gosse Bouma back in 1988, in the framework of categorial unification grammar [Bouma 1988].

As a closely related approach we should also mention here the analysis of [Maier and Steffens 1988] which is likewise embedded in a categorial unification grammar framework.

terminals taking the position and function of the head noun. In section 4.6 we argue that the assumption of an empty terminal is motivated not simply by theory-internal constraints, but also to a large degree by the empirical data. We then explore different solutions to the problem that empty nominal heads can occur only in a certain syntactic environment. Specifically, there are constraints on the left periphery of nominal phrases which are not easy to capture given that determiners, adjectives and nouns are all optional constituents.

Section 4.7 will be concerned with the morpho-syntactic phenomena of agreement and the distribution of the declension paradigms of specifiers, adjectives and nouns in German. We analyze the declension class in nominal phrases (strong or weak) as a property which should be described on the level of `MINOR` features. Declension classes in our approach are not based on the inflectional paradigms of individual determiners, but rather on general inflectional properties which cut across these paradigms. Under this assumption, the declensional variation in nominal phrases can be treated as an instance of government rather than agreement, with the determiner governing the declension class of its complement. The fact that the description of the entire declension system then reduces to four simple and very general axioms can be taken as further evidence for the treatment of determiners as functional heads. This analysis of declension phenomena is taken one step further in section 4.8 where we explore the specific problem of numerals potentially co-occuring with full-fledged determiners. We provide an analysis which takes into account the intermediate status of numerals sharing functions with both determiners and adjectives.

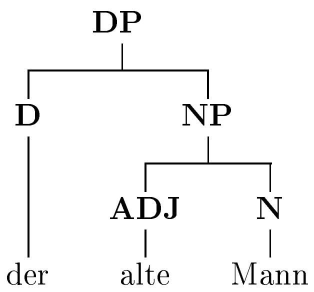
In the last section 4.9 we compare our analysis of the three classes of phenomena described—determinerless constructions, declension and adjectival adjunction—with the approach suggested in [Pollard and Sag 1994]. We show that their approach to specifiers, which is based on a mutual selection of nouns and determiners, requires disjunctive subcategorization for the treatment of determinerless constructions, where our analysis is based on underspecification. With respect to declension, we argue that we can describe apparent idiosyncratic exceptions in a more straightforward and general way, while their approach involves some stipulations which have to be stated very indirectly on selected forms in the otherwise fully regular adjectival paradigm. Finally, we show that the approach of [Pollard and Sag 1994] overgenerate in the case of adjectival adjunction, since it does not provide a sufficient basis for the distinction between maximal and non-maximal nominal projections. We will suggest an alternative solution to the problem, consistent with [Pollard and Sag 1994], which is based on a generalization of the so-called *head-marker schema* and an extension of the Specifier Principle. Given this modification, our approach and the one by [Pollard and Sag 1994] have almost the same weak generative capacity. We nevertheless maintain that the analysis of [Pollard and Sag 1994] is considerably more complex.

4.2 Heads of Nominal Phrases

In determining the category of nominal phrases, most traditional approaches choose the noun to function as the head of the phrase. One reason for this assumption is presumably that categories corresponding to maximal projections are traditionally defined through the "meaningful" major, substantive or lexical categories, such as noun, verb, adjective or preposition, rather than through other, minor or functional categories, such as determiners or complementizers.

However, approaches which treat functional categories as the heads of phrases and maximal projections have become increasingly popular in recent years, especially in the tradition of the Government and Binding framework (G&B). This concerns not only overt and non-overt functional categories, such as COMP or C and INFL or I, but also the treatment of the determiner as the head of a nominal phrase, the so-called DP-analysis.[2] Ignoring possible variants and irrelevant details, the structure assigned to a nominal phrase in a DP-analysis would be something like (96), where the determiner functions as the head of the maximal nominal projection while the noun phrase is interpreted as a complement or argument to the determiner.

(96)



In the following section we will discuss the consequences and difficulties of pure NP- and DP-analyses, in which the noun and the determiner respectively are interpreted as heads of maximal nominal projections. Ignoring a number of other issues, we will concentrate here on the problem of determinerless constructions and the question of appropriate head features involving in particular the phenomenon of declension. Our main emphasis will also be on implications of these approaches given the assumptions on heads made in the HPSG framework.

4.2.1 Headless Constructions

The traditional definition of endocentric constructions as those in which one constituent determines the category of the resulting phrase does not give us any

[2] Cf. among others [Abney 1987], [Haider 1988], [Stowell 1989].

guidance, since it is exactly the question what category should be assigned to nominal phrases which we have to decide on. However, in this definition there is also a hidden assumption that the one constituent which has to be obligatorily present in order to form the phrase should be the head. Thus, one way to approach the choice between nouns and determiners as heads of nominal phrases is to examine what happens if one of these potential heads is missing and therefore cannot overtly provide the appropriate head features to determine the category of the maximal projection. With the assumption of empty heads the range of possible analyses becomes of course less constrained, but there are also a number of other theoretical and empirical restrictions which can be taken into account.

Determinerless Constructions

Constructions with a missing determiner typically involve either plural nouns (e.g. *Kinder*) or mass nouns (e.g. *Wasser*). These constructions qualify as 'maximal' projections of a nominal phrase even though they are missing a specifier.

In an *NP-analysis* the solution to the problem of missing determiners very much depends on whether there is a lexically determined relationship between the determiner and the noun, or whether this configuration is determined on purely structural grounds. If we assumed that there is no lexical relationship in the sense of the noun selecting or licencing the determiner, the only possibility for the derivation of determined and determinerless constructions appears to be the postulation of two different phrase structure schemata. One of these schemata would be unrestricted with respect to the nominal head and would obligatorily introduce a determiner. Another rule would describe the unary, determinerless expansions of the maximal category NP and would have to be constrained to nominal heads which are not singular count nouns. Clearly, such an approach would run very much counter to the spirit of HPSG, since it would require the introduction of additional rule schemata which must even be sensitive to specific categories.

Thus, the most obvious approach within HPSG is to postulate lexical selection of the determiner through the noun. However, unless one introduced a new selection mechanism, the only appropriate means to encode this selection is subcategorization, which can require the presence or absence of individual elements. The treatment of the presence or absence of determiners in this analysis accordingly boils down to the treatment of optional complements. Singular count nouns obligatorily subcategorize for a determiner. All other forms would be specified for an optional determiner, which can be equivalent to introducing disjunctive subcategorization frames.

Both the purely structural and the lexical approach could of course be "unified" if one adopted a Uniform Constituent Hypothesis, stipulating that all phrasal configurations of a certain kind have to follow the same pattern or tree geometry at any cost (say the one suggested by $\overline{\text{X}}$-syntax). Under such an assumption those constituents which do not correspond to an overt string are simply represented as an empty terminal. In the G&B framework there appear to be very few inhibitions to adopt such a solution, however the question arises what the predictive power of such an assumption is.[3]

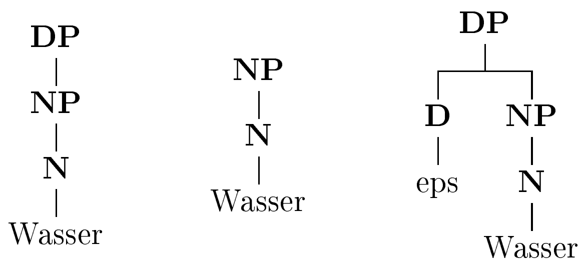
In a *DP-analysis* the problem of determinerless constructions turns out to be even more critical, since here the designated head of the nominal phrase is missing. At first glance there seem to be three different solutions:

(a) We stipulate a unary rule which expands a DP exclusively into an NP.

(b) We assume disjunctive subcategorizations on the matrix predicate for either a DP or an NP.

(c) We postulate an empty determiner.

(97) (a) (b) (c)



Option (a), which turns the noun into the head of a DP if the determiner is missing, we can exclude in this form straightaway if we follow the standard HPSG assumptions. As long as we want to maintain the HFP, we cannot stipulate a phrase structure such as (98), since there is no token identity between the `HEAD` features of the mother and the `HEAD-DTR`.[4]

[3]See also the discussion of [Kornai and Pullum 1990] below.

[4]For expository reasons we assume here that the category or part-of-speech information is encoded as an atomic value of a `HEAD` feature `MAJOR`, e.g. [`MAJOR` *N*] or [`MAJOR` *D*]. This way of categorizing parts of speech has been replaced in [Pollard and Sag 1994] by the use of types which are defined for the respectively appropriate `HEAD` features. Of course, if one assumed that the two types *noun* and *det* were not disjoint but rather stand in a subtype relation, such a unary rule would not necessarily lead to a unification failure anymore.

(98) $\begin{bmatrix} \texttt{SYNSEM} \begin{bmatrix} \texttt{LOC}|\texttt{CAT}|\texttt{HEAD} \begin{bmatrix} \texttt{MAJOR}\ D \end{bmatrix} \end{bmatrix} \\ \texttt{DTRS} \begin{bmatrix} \texttt{HEAD-DTR}|\texttt{SYNSEM}|\texttt{LOC}|\texttt{CAT}|\texttt{HEAD} \begin{bmatrix} \texttt{MAJOR}\ N \end{bmatrix} \end{bmatrix} \end{bmatrix}$

Note, however, that this analysis would not be excluded in a framework such as GPSG, where the Head Feature Convention is considered a *default* principle and where values of head features can be overwritten.[5] Even if our own analysis below is somewhat related to option (a), it turns out to be theory consistent as well as to avoid many other undesirable aspects, such as unary rules.

Option (b) is based on the assumption that if a determiner is present it defines the categorial properties of the maximal projection as DP. If it is missing, the noun will become the head and project into a different category, NP. This solution allows to maintain the HFP within the nominal phrase, at the price of shifting the disjunctive specification to the subcategorization of the selecting matrix lexeme. One would have to stipulate that *every* lexeme taking a nominal complement accepts either a DP or an NP. A solution, which is comparable to this approach, is adopted in HPSG for the analysis of verbal or clausal complements. Rather than analyzing all verbal clausal projections as the same category S, there is a distinction between non-finite VP constituents (missing a subject) and full S projections including an overt subject phrase.

For the analysis of nominal phrases, such a solution is linguistically unsatisfactory, because among other reasons it implies that the categorial properties of a maximal nominal phrase and the selectional restrictions of a matrix lexeme vary as a function of properties such as `NUMBER` or mass vs. count noun, which are predominantly of phrase-internal relevance.[6]

Version (c), which would be a good candidate for a standard G&B solution, allows us to maintain the HFP along with a homogeneous subcategorization on the matrix lexemes. However, it systematically requires the introduction of a phonetically empty element of the category Determiner which subcategorizes for an NP. Again this is unsatisfactory, since the only motivation for such an empty determiner would be to provide a means of 'recoding' or transforming the phrasal

[5]Cf. [Pullum 1991] who exploits this fact for the redefinition of categories.

[6]John Nerbonne pointed out to us that if this were an adequate solution one would at very least expect to find lexemes which select exclusively either an NP or DP. However, the only cases that we are aware of that might come close to this behaviour are copula constructions like (i) where the predicative is a determinerless count noun.

(i) Hans ist Lehrer.

Since in the case of verbal complements we do find such differences in selection, it appears that a disjunctive subcategorization can be defended for these constructions.

category headed by the noun, and to licence the presence of the NP through its subcategorization frame.

Again, this would certainly be a viable solution under a Uniform Constituent Hypothesis. At first sight, such an assumption makes the theory appear more uniform and also constrained. Yet, under closer scrutiny, such an approach is nothing more than a slightly more elegant and consistent reformulation of solution (a), the direct recoding and transformation of a category by means of a unary rule. In the G&B tradition this problem and solution keep recurring, another instance being the empty complementizers which are necessary if all sentences including root sentences are analyzed as CP, i.e., maximal projections of complementizers. One of the most important arguments against such an approach is provided by [Kornai and Pullum 1990][39ff] who show that a non-monotonic or default head feature convention is equivalent to a framework obeying a monotonic head feature principle, but allowing for the introduction of empty heads. In the extreme case the implications of this proof are that what appears to be a restriction on possible phrase structure configurations at first sight, turns out to provide no restriction whatsoever.

Empty Head Nouns

While determinerless constructions are a very regular and systematic phenomenon, there is also the counterpart of missing nouns in constructions which are equivalent to English *one*-anaphora constructions. The adequate treatment of this phenomenon could provide an argument against the postulation of determiners as head or in general against the postulation of empty determiners.

The regularity that has to be accounted for is that in German all of the structural positions at the left periphery of a nominal phrase—i.e., the positions of the specifier, adjectives and noun—are in principle optional. However, there is a constraint that at least one of these positions has to be filled. All of the expressions in (99) except for the last one are grammatical nominal phrases in some context.

(99)	die	alten	Männer	mit Hut
	die	alten	–	mit Hut
	die	–	Männer	mit Hut
	die	–	–	mit Hut
	–	alte	Männer	mit Hut
	–	alte	–	mit Hut
	–	–	Männer	mit Hut
	*–	–	–	mit Hut

If we assumed that missing determiners are neither treated as empty heads (in a DP-analysis) nor as empty terminals (in an NP-analysis), there is a comparatively simple way to capture these restrictions, namely to require that the left-most constituent in a nominal phrase must not be an empty terminal.

If we assumed that the noun is the head and the missing determiner is represented by an empty terminal, the constraint would be considerably more complicated and would have to take into account the category of the left-most empty terminal. If we assumed that the determiner functions as head, one would have to stipulate that the restriction on the non-empty left periphery does not apply to maximal nominal phrases but only to the complements of empty determiners. Configurations such as (100) could be excluded if the empty determiner (through its subcategorization) required its complement to have such a non-empty left periphery.

(100) * [$_{DP}$ *eps-d* [$_{NP}$ *eps-n* mit Hut]]

To be precise, it would have to be stipulated that the left-periphery of the complement of an empty determiner must not be filled by an empty noun, rather than by some empty element in general. Since the restriction holds because constructions such as (101) are by all means grammatical, which in this approach would have to be analyzed as having an empty determiner whose complement-NP begins with an empty determiner itself.

(101) [$_{DP}$ *eps-d* [$_{NP}$ [$_{AP}$ [$_{NP}$ *eps-d* Bücher] lesende] Kinder]]

This approach could have the advantage that the relevant constraints can be captured on a lexical level, albeit at the price of stipulating two instead of only one (or even no) empty terminal category.

We will discuss these questions in considerably more detail below. However, the problem and its various solutions already illustrate the potential range of motivations for an empty node in a linguistic theory. On the one hand, there is the case of empty nouns, which—as we will show—can be motivated and appear to be necessary for anaphoric interpretation; on the other hand, we have empty determiners which are postulated purely for the sake of maintaining a uniform structural representation. While the former could be motivated even across different linguistic theories, the latter appears to be based exclusively on theory specific assumptions for which there is hardly any other motivation than uniformity of structural representations.[7]

[7] The postulation of an empty head noun in these constructions is well motivated by, among other reasons, the fact that the empty nouns have the function of genuine anaphora with a distribution very similar to English *one*-anaphora.

This issue is also taken up by [Pullum 1991][776] who postulates that empty terminals should be motivated by extraction sites, by anaphoric links to other constituents (PRO) or by an independently identifiable semantic content. In a maximally constrained HPSG, the first two cases are presumably obsolete or at least can be eliminated in any event. The last case, which Pullum illustrates by constructions such as [$_{NP}$ *the rich* [$_{N}$ *e*]], in fact almost coincides with our postulation of empty noun anaphora.

4.2.2 Head Features

As we already discussed above, the head of a phrase is recursively defined in HPSG, as in many other theories, as the category from which the mother node inherits a specific set of properties: the head features. Through postulates such as the Head Feature Principle (HFP), these properties are percolated from the lexical head through intermediate projections to a maximal projection. As we argued, the choice of the set of relevant head features should be motivated either by external relations—such as agreement or selection—which may hold between the maximal projection and another category, or by internal relations crucially involving intermediate projections.

Externally relevant head features

In German, it is quite obvious that all relevant morpho-syntactic features which must be projected to the maximal projection of a nominal phrase for the purpose of external agreement or government manifest themselves equally well on the noun or the determiner itself. These features comprise above all `CASE`, which can be governed by external heads, and the features `NUMBER` and `PERSON` which are relevant, in German, to the agreement constraints in relations between the subject and the finite verb. In addition `GENDER` will play a role if anaphoric relations (including relative pronouns) are concerned.

As far as the features `CASE`, `NUMBER` and `GENDER` are concerned, we find a high degree of syncretism or morpho-syntactic ambiguity in the forms of the determiners and nouns, however, it would be unjustified to claim that either of these categories provides more of this information than the other.

One of the few cases which could be induced as exceptions to this generalization are expressions like (102), in which the pronoun could be analyzed as a head in a determiner position subcategorizing for the noun.

(102) $\text{Du}_{2\text{nd}}$ $\text{Idiot}_{3\text{rd}}$ $\text{hast}_{2\text{nd}}$ die Milch vergessen.

If we accept such an analysis and interpret subject-verb agreement as a syntactic phenomenon, these examples provide evidence that agreement is formulated more adequately through the `PERSON` feature of the determiner than that of the noun, and that the determiner should therefore function as the head of a nominal phrase.[8]

However, it has to be kept in mind that at least for the feature `PERSON` it may be equally adequate to assume a semantically based account for agreement. Morpho-syntactically relevant distinctions with respect to `PERSON` manifest themselves

[8]Gregor Erbach and Reinhard Karger brought this class of examples to our attention.

only on finite verb forms and on personal (or reflexive) pronouns. The question is therefore whether non-pronominal noun phrases should be syntactically marked for the attribute `PERSON` at all. One of the interesting cases corroborating this view is relative pronoun agreement in German.

(103) ihr$_{2^{nd}}$, die$_{?}$ das nicht wißt$_{2^{nd}}$ / *wissen$_{3^{rd}}$, ...

(104) du$_{2^{nd}}$, die$_{?}$ das nicht gelesen hast$_{2^{nd}}$ / *hat$_{3^{rd}}$, ...

What we find here is that the relative pronoun does not show any `PERSON` agreement, but only `NUMBER` and `GENDER` agreement with the phrase to which the relative clause is attributed. However, the finite verb has to agree with respect to `PERSON` with this head noun rather than the relative pronoun. Of course, one could argue that relative pronouns do not bear any `PERSON` agreement at all. In the case of free relative clauses (105) the third person marking on finite verbs would then have to be induced by some default mechanism.[9]

(105) Die$_{?}$ das nicht wissen$_{3^{rd}}$, sollen aufstehen.

This means that we have the choice of either stipulating a syntactic default or of assuming that the morpho-syntactic person marking of the finite verb is inferred from semantic properties. There is little evidence why non-pronominal noun phrases, clausal and 'implicit' subjects should not be marked as *non-speaker* and *non-addressee* on a semantic level. This would provide us with the necessary information for determining the form of the finite verb, as long as we can identify what semantic argument stands in a semantic agreement relation with the finite verb.

A similar argument as for `PERSON` as a predominantly semantic property has been made with respect to `NUMBER` and also `GENDER` agreement. However, while we are willing to subscribe to the former view, we consider the evidence for analyzing `NUMBER` as a semantic relation as much weaker and practically non-existent in German if it comes to `GENDER`.

To summarize, among the head features relevant for external selection of noun phrases, we primarily count `CASE`, `NUMBER` (and to a limited degree also `GENDER`), while we admit that relations involving `PERSON` could be treated more appropriately on a semantic level. Since the relevant morpho-syntactic distinctions manifest themselves equally well in the morphological paradigms of determiners,

[9] The effects of such a 'default' can be also observed in other constructions which lack a nominal (or rather nominative) subject, i.e. sentences with a clausal subject (i) and sentences without an explicit subject (ii):

(i) Daß du nicht kommst, stört$_{3^{rd}}$ mich.

(ii) Gestern wurde$_{3^{rd}}$ gearbeitet.

adjectives and nouns, we do not see an argument in favour of choosing determiners or nouns as heads. In our analysis below, we show how a syntactic agreement relation even for the feature `PERSON` can be maintained for the cases discussed above, if the determiner is analyzed as a head.

Internally relevant head features

The head features which are involved in relationships internal to a nominal phrase are the features `NUMBER`, `CASE`, `GENDER` and `DECLENSION`.

Among these features, `NUMBER` and `CASE` do not provide much evidence either for deciding between nouns and determiners as heads either. The feature `GENDER` is a property which (in most cases) is inherently lexically determined by the noun, but has a morphological reflex on determiners and adjectives. Since in a unification-based approach to agreement all these features would be shared by both the noun and determiner, there is no reason for preferring either as the appropriate head.

However, there is one internally relevant property for which determiners, adjectives and nouns may have values which systematically differ in a crucial way, namely `DECLENSION`.

As we pointed out above, the most straightforward way to implement an NP or DP-analysis in HPSG is to let the noun or the determiner respectively select the other category. Since selection to some degree also entails government, the question is whether the morpho-syntactic relationships between nouns and determiners are more appropriately defined as the noun governing the determiner or the other way around. One rather complex phenomenon which can be interpreted as evidence against the former analysis is the declension of adjectives, determiners and nouns in German. Since the discussion of all the ramifications of this problem would take us too far afield and since we will come back to this issue below in much more detail, we illustrate this problem here by focussing on just some few aspects.

Inflected adjectival forms and, as can be argued, determiners, numerals, and nouns, come in two different declension classes: a weak and a strong paradigm.[10] What form of declension an adjective or noun takes is largely determined by the presence or absence of a determiner as well as the determiner's inflection. The simplest instances of these regularities are illustrated by the opposition between determined and undetermined plural forms, where we find weak inflection of the adjective and noun in (106) and strong inflection in (107), respectively.

(106) meine$_{st}$ lieben$_{we}$ Verwandten$_{we}$

[10] In the analysis given below the third, *mixed*, declension class becomes obsolete.

(107) liebe_{st} Verwandte_{st}

The situation is slightly more complicated in the case of inflected numerals, which can optionally occur together with determiners. If such a numeral is preceded by a strong determiner it has to be weakly inflected (108)/(109), otherwise it is strong itself (110). In all cases a following adjective must be weak.[11]

(108) meine_{st} beiden_{we} lieben_{we} Verwandten_{we}

(109) meine_{st} lieben_{we} beiden_{we} Verwandten_{we}

(110) beide_{st} lieben_{we} Verwandten_{we}

In a DP-analysis these phenomena can be captured in a straightforward way by claiming that determiners and numerals function as heads and that the highest overt head in a nominal phrase, be it determiner, numeral or noun, must carry strong inflection, that it governs the following constituent and that it requires it to be weakly inflected. As a necessary further refinement we would assume that uninflected determiners (*mein* in (111)) and numerals (*zwei* in (112)) simply pass on their declension class to the governed constituent.

(111) mein_0 liebster_{st} Verwandter_{st}

(112) meine_{st} liebsten_{we} zwei_0 Verwandten_{we}

In contrast, an NP-analysis will have to determine the prenominal variation of declension classes by making disjunctive subcategorizations galore. If the noun functions as the head and selects the determiner, the required declension of determiner would have to be specified on the selecting head noun. In order to cover the numeral construction in (110)/(112) one must either assume that the category of numerals is ambiguous between determiner and adjective, or stipulate that a weak noun allows either a numeral or a determiner as a complement.[12] If one provided a third subcategorization frame accounting for the presence of both determiner and numeral, one would have to stipulate that the determiner is marked strong, while the numeral is marked weak, and so on.

As we will show in more detail below, a pure NP-analysis may also be forced to take recourse to rather indirect encoding strategies to cover this phenomenon

[11] We follow the observation in [Drosdowski 1984][293] that the strong inflection after *beide* is becoming obsolete.

[12] If the analysis of numerals as adjectives were correct, it would be hard to explain why constructions such as (ii) are ungrammatical.

(i) die zwei alten Männer

(ii) *die zwei und alten Männer

Though coordination may not be a reliable test of categorial similarity, we are not aware of cases where two constituents of the same category cannot be coordinated.

running also against certain locality constraints, while the generalizations can be expressed quite straightforwardly in an approach analyzing the determiner as the head of a nominal phrase.

4.2.3 Adjunction and Submaximal Projections

So far we only considered the question of what category should provide the head of a nominal phrase as a maximal projection. However, this question can also be related to the problem of defining submaximal levels of projections as they are required for pre-nominal adjunction of attributive adjectives. Pre-nominal adjunction is more critical than post-nominal adjunction, since it will be crucial that attributive adjectives adjoin before the determiner is added, while it can only be decided on theoretical grounds at what level post-nominal adjuncts are added. For post-nominal adjunctions it suffices to impose the constraint that post-nominal genitive phrases have to combine with nouns at the lowest, lexical level of projections, while other adjuncts can adjoin at any higher level of projection.

Thus, to capture a phenomenon like pre-nominal adjunction it appears indispensable to define an intermediate level of phrase structure projection, which is below the level of maximal projections. For the analyses referred to above this means that adjectives have to be adjoined either to the NP node (or a lower projection of the noun) in a DP-analysis, or to an intermediate projection $\overline{\mathrm{N}}$ in an NP-analysis. In both cases, this assumption is enforced by the fact that there must be a structural position higher than the adjectival adjunction site, which can be reserved to the determiner to ensure the appropriate linear order of the determiner preceding the adjectival adjuncts. This also means that one is forced in both cases to assume additional structure, if the determiner is not present. In a DP-analysis this additional structure has to be represented in the form of an empty determiner, which may be necessary for independent reasons anyway. In an NP-analysis the additional structure has to be represented at least in the form of a unary expansion of the maximal NP node to the submaximal adjunction node.

In a pure phrase-structure-based approach, such additional structure may not appear to be too unusual. However, under closer inspection the necessity to introduce potentially redundant or vacuous structure turns out to be an effect of the assumption of a uniform constituency again. Since pure phrase structure approaches foresee only node labels and dominance relations as a means of encoding grammatical relations, there is simply no way to reconcile the needs of defining adjunction sites and maximal projections other than by tree configurations.

However, while HPSG clearly does provide other means for describing phrase structure than just node labels, it is interesting to see that a pure NP-analysis will nevertheless enforce such vacuous structure to be built in order to prevent overgeneration. Although feature structures as a formal device would be ideal to generalize over the different levels of projection by means of underspecification, even in HPSG an NP-analysis appears to be viable only if some empty determiner or some unary derivation is assumed. Why this is the case we will describe in more detail below, where we also demonstrate that the traditional assumptions of HPSG may even be insufficient to block overgeneration of constructions like **red the roses* to start with. Again our own analysis of functional heads will naturally lend itself to control these phenomena by means of underspecification and without the postulation of such vacuous structure.

4.2.4 Summary

From the few issues we have presented so far in the comparison of pure DP- and NP-analyses, it seems that, if there should be any strong arguments in favour of the latter, they can be derived mainly from the problem of determinerless constructions. A pure DP-analysis for determinerless constructions is either untenable through violation of well-motivated restrictions of the theory (the HFP), or based on unsatisfactory linguistic assumptions: disjunctive subcategorizations or the stipulation of empty heads for the purpose of recoding the category of a maximal nominal projection. However, the problem of defining submaximal projections may enforce the postulation of empty determiners or unary expansions even in an NP-analysis. With respect to empty nominal heads both approaches could be expected to be almost equally complex. With respect to the determination of relations such as government of declension, a DP-analysis is likely to turn out more elegant and straightforward. Still, neither of the two approaches appears fully satisfactory if pursued in its purest form.

In the next section we therefore present an analysis of specifiers or functional heads which allows us to maintain a DP-oriented analysis, while at the same time dispensing with the need for empty determiners and other vacuous structure. This analysis will build mainly on the formal device of underspecification which we consider to be more appropriate to a heavily constraint-oriented framework like HPSG, in which traditional phrase structures should play a much more subordinate role.

4.3 Substantive and Functional Categories

In the following, we outline the categorial basis of our analysis. We describe the notions of *substantive* and *functional* categories and how this distinction can be reflected in the geometry of HEAD features. On this basis, we define a primitive notion of *functional completeness*, which becomes relevant for the definition of *maximal projections*. These abstract definitions are applied to the analysis of determined and determinerless nominal constructions in the next sections and to the analysis of complementizers and finite verb position in chapter 5.

In our analysis, we take up the traditional distinction between substantive (or lexical) and functional categories or major and minor categories.[13] Typically, categories such as nouns, verbs, adjectives and adverbs have been subsumed by the former terms, whereas the latter class is taken to subsume amongst others determiners, quantifiers, complementizers and sometimes also prepositions.

In standard HPSG, the distinction between major and minor categories is assumed to be reflected as a predominantly syntactic type distinction over the possible values of the attribute HEAD [Pollard and Sag 1994][22ff], which are partitioned into two types *substantive* and *functional*. Most of the subtypes of these two types are defined for specific attributes to be appropriate, such that most types of categories (or parts of speech) are defined for a disjoint subset of HEAD features. Apart from this distinction there is no effective constraint as to whether functional categories could function as selectors or not. However, it is quite obvious that only one category in a phrase can function as the head of a phrase and it is standardly assumed that it is the substantive category which provides the head features.

In contrast to this approach, Lexical Functional Grammar foresees that the distinction between major and minor categories is captured by major categories bearing a (globally) unique *semantic form*, which is encoded as the value of a PRED feature and which expresses the subcategorization or selection constraints of a head. This attribute is missing on minor categories, which thus never select any other phrase. The effects of this assumption are on the one hand that a phrase cannot be headed by two major categories, since the two PRED values would cause a conflict. On the other hand, a phrase can be 'headed' by a major and a minor category simultaneously in the sense that the information associated with the minor and the major category is identified with the phrase's root node. Well-formedness restrictions require that every maximal projection must contain

[13]These categories could also be referred to as content words and function words or open class and closed class lexical items. However, this terminology appears to subsume a more heterogeneous range of categories under the latter terms, such as all kinds of particles or adverbs, for which we do not want to claim the same status as for the functional categories we have in mind.

a semantic form and thus exclude configurations which are exclusively headed by a minor category.

Our approach basically follows most of the HPSG assumptions, however it also incorporates some of the ideas from LFG. As in standard HPSG we assume that categories or parts of speech are predominantly characterized through head features, which may be partitioned into different subclasses. However, we do not assume that this partitioning also encompasses a distinction between functional and substantive properties, but rather assume, similar to LFG, that a phrasal sign can be defined for subsets of both properties simultaneously. This means, above all, that we do not make a prediction at the level of a phrase as to whether the head of that phrase must be a functional or a substantive category. Specifically, we assume that substantive and functional properties of categories are encoded by means of two `HEAD` features, `MAJOR` and `MINOR`, respectively. The most abstract type of categories is given in (113), which represents the type of the `SYNSEM` value of the type *sign*.[14]

(113) $$\begin{bmatrix} \texttt{LOC} & \begin{bmatrix} \texttt{CAT} & \begin{bmatrix} \texttt{HEAD} & \begin{bmatrix} \texttt{MAJOR} & [\ ] \\ \texttt{MINOR} & [\ ] \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}_{cat}$$

For the time being we assume that categories will have other `HEAD` features besides the attributes `MAJOR` and `MINOR`, as for example the feature `MOD`. However, in the long run one could envisage to strengthen the feature geometry so that all `HEAD` features are distributed between the two classes of attributes. While we give a more general specification of the possible values of the two features `MAJOR` and `MINOR` in this section, we will concentrate on nominal (and adjectival) categories throughout the rest of this chapter and will come back to verbal categories in the next.

4.3.1 Major Categories

For lexical substantive categories, we assume that they contribute a basic part-of-speech information (`POS`) in the specification of the `MAJOR` attribute. For the definitions of the part-of-speech of the major categories noun, adjective, verb and preposition/adverb, we follow the traditional representation in terms of two binary features [`N` ±] and [`V` ±]. The cross classification of these features gives us the arrangement (114) for the `POS` values of major categories.

[14]In the following we introduce type definitions for categories at the level of the value of `SYNSEM` since the distinction between lexical and phrasal signs plays at best a secondary role in our analysis. Feature structures in tree diagrams represent the `CAT` value of a node.

(114)

	Noun	Adj	Verb	Prep
N	+	+	−	−
V	−	+	+	−

Factoring the description of the major categories in this way allows us to generalize over natural classes of categories by means of underspecification of some value.

As already indicated above, this underspecification can be exploited, for example, in the specification of LP constraints. Whereas, for example, in English and French one assumption is that all lexical heads can be constrained to a left-peripheral position, one could assume a differentiation for German which runs along the distinction encoded by the feature V. Categories which are characterized as [V +], i.e., adjectives and verbs, typically occur in a right-peripheral position (relative to case-bearing arguments and adjuncts), whereas the categories marked as [V −] (with the exception of a few postpositions) have to be left-peripheral.

Besides the part-of-speech information the MAJOR attribute of every category can be defined for a set of further attributes which encode the appropriate properties of that category.

For our purposes here, we assume a type *nominal* which generalizes over the categories adjective and noun in the obvious way. This type is defined for an additional complex-valued MAJOR feature AGR which comprises most of the features that are relevant for relations internal to nominal phrases, i.e., NUMBER, GENDER and CASE in (115), excluding the feature DECL, however, which encodes the declension class of nominal categories.

(115)
$$
{}_{agr}\begin{bmatrix}
\texttt{NUM} & {}_{number}[\,] \\
\texttt{GEN} & {}_{gender}[\,] \\
\texttt{CASE} & {}_{case}\begin{bmatrix} \texttt{GOV} & bool \\ \texttt{OBL} & bool \end{bmatrix}
\end{bmatrix}
$$

For the value of the feature CASE, we assume a type *case* which stands for the combination of two boolean features OBL and GOV as given in (116). This split again accounts for underspecification and generalization over individual cases.[15]

(116)

	Nom	Acc	Dat	Gen
GOV	−	+	+	−
OBL	−	−	+	+

The type *nominal* is then defined as in (117).

[15]To our knowledge, a classification along these lines was first suggested by Manfred Bierwisch.

(117)

$$nominal \equiv cat \wedge \left[\texttt{LOC|CAT|HEAD|MAJOR}\begin{bmatrix}\texttt{POS} & \begin{bmatrix}\texttt{N} & +\end{bmatrix}\\ \texttt{AGR} & agr\begin{bmatrix}\ \end{bmatrix}\end{bmatrix}\right]$$

The MAJOR attributes of verbal categories in German always come in two distinct forms depending on whether the category is marked as finite or non-finite. Thus, the general type *verb*, introducing the part-of-speech information, is partitioned into two subtypes *fin-verb* and *non-fin-verb*.

(118)

$$verb \equiv cat \wedge \left[\texttt{LOC|CAT|HEAD|MAJOR}\begin{bmatrix}\texttt{POS}\begin{bmatrix}\texttt{N} & -\\ \texttt{V} & +\end{bmatrix}\end{bmatrix}\right]$$

$$\mathrm{part}(verb) \equiv \textit{fin-verb}, \textit{non-fin-verb}$$

Finite verbs we assume to be defined through the presence of an agreement feature comprising NUMBER and PERSON.[16] This type represents the supertype for both tensed and imperative forms.

(119) $$\textit{fin-verb} \equiv verb \wedge \left[\texttt{LOC|CAT|HEAD|MAJOR}\begin{bmatrix}\texttt{AGR}\begin{bmatrix}\texttt{NUM} & [\ ]\\ \texttt{PERS} & [\ ]\end{bmatrix}\end{bmatrix}\right]$$

The distinction between tensed verbs and imperative forms of verbs is then captured by the former being defined for two binary features TENSE and MOOD with the possible values *pres, past* and *ind, conj* respectively, whereas the latter are undefined for these features.

(120) $$\textit{tensed-verb} \equiv \textit{fin-verb} \wedge \left[\texttt{LOC|CAT|HEAD|MAJOR}\begin{bmatrix}\texttt{TENSE} & tense\begin{bmatrix}\ \end{bmatrix}\\ \texttt{MOOD} & mood\begin{bmatrix}\ \end{bmatrix}\end{bmatrix}\right]$$

(121) $$\textit{imper-verb} \equiv \textit{fin-verb} \wedge \left[\texttt{LOC|CAT|HEAD|MAJOR}\begin{bmatrix}\texttt{TENSE} & *undef*\\ \texttt{MOOD} & *undef*\end{bmatrix}\right]$$

Non-finite forms we assume to be divided into the forms of the supinum and participle by a boolean feature PART, the past participle and infinitival forms by a boolean feature PERF and the bare and *zu*-infinitive by an attribute ZU. The supertype of all non-finite forms is then specified as (122).[17]

[16] For the sake of simplicity, we ignore here the discussion above suggesting that these two properties could also be treated on a semantic level.

[17] The distinctions we assume for non-finite forms basically follow the classification suggested by [Bech 1983].

$$(122)\quad \textit{non-fin-verb} \;\equiv\; \textit{verb} \;\wedge\; \left[\texttt{LOC}|\texttt{CAT}|\texttt{HEAD}|\texttt{MAJOR}\begin{bmatrix}\texttt{PART} & \textit{bool}\\ \texttt{PERF} & \textit{bool}\\ \texttt{ZU} & \textit{bool}\end{bmatrix}\right]$$

For prepositional categories, i.e., prepositions, proper adverbs (*leider*), or pronominal adverbs (*darauf*), we only assume that they are marked by the POS attribute. The only MAJOR feature, which may be necessary in addition is PFORM which encodes the form of the (underlying) preposition for the purpose of prepositional object selection.

Next to these category specific attributes, we take it that there will be a number of attributes which cut across categories, such as the attribute MOD defining adjunctive categories, or a boolean attribute PRD distinguishing predicative and non-predicative categories.

It should be noted that the attribute POS encoding the parts of speech could be equally well encoded by partitioning the appropriate sub-types of the type *cat*. For the moment, we keep this attribute mainly for transparency reasons.

4.3.2 Functional Categories

In the following we outline a notion of functional categories that attempts to capture properties that we consider typical for lexically expressed minor categories such as determiners or complementizers. However, we also envisage that this notion can be extended to subsume specific morphological and structural devices. For example, it appears tempting to consider different forms for the expression of comparatives, such as the lexical marker *more* and the morphological inflection *-er* in English, as two different instances of the same functional category. Similarly, the complementary distribution of complementizers and V-initial construction in German can be regarded as arising from lexical and structural realizations of the same property. We will discuss this latter extension of the notion of functional category at length below in chapter 5.

The attempt to cover such a broad range of devices under the notion of functional categories will of course lead to a considerable amount of variation within the class, which will make it sometimes difficult to draw a clear-cut boundary between substantive and functional categories. The following descriptive characterization of functional categories should therefore be seen as a set of prototypical properties.

At least in terms of their syntactic function and with respect to the set of lexical items realizing them, functional categories typically form closed classes of lexical items. For example, even if we extend the notion of complementizers to include adverbial subordinative conjunctions, such as *als*, *nachdem* etc., we still end up with a fairly small set of items.

Functional categories clearly cannot be regarded as entirely void of semantic information if we subsume determiners and quantifiers under this class. However, their roles appear to be predominantly of a syntactic nature or to be quite different from the semantic types of open class lexical items. The German complementizer *daß* is an instance of a prototypical functional item, which can hardly be assigned any meaning at all.

Functional categories typically do not occur on their own but must combine with a substantive category. Again, the prime example would be complementizers which are bound to the presence of a clause and which outside a contrastive context do not even allow an ellipsis of this substantive category. Determiners are more variable in this respect, since they may occur in isolation. However, the typical analysis for these constructions either assume that the determiner has undergone some (pro-)nominalization or that they are accompanied by an empty noun (as we suggest below).

However, in contrast to major categories, which may take any number of complements, functional categories normally combine with only one major category. Typically this major category must be fully saturated.[18] This generalization goes together with the observation (or even implies) that functional categories attach in a position which is peripheral relative to the major category and its complements.

In comparison with major categories, which normally can subcategorize for all different types of other categories, functional categories are much more selective with respect to the major categories they combine with. Whereas verbs, for example, can select the whole range of substantive categories, complementizers only occur with clauses or verbal projections, and determiners are typically restricted to the combination with nominal categories. [Baltin 1989][6] goes even as far as to make this a defining property of functional categories:

> A MINOR LEXICAL CATEGORY is a category all of whose members have the same subcategorization frame with respect to grammatical categories.

It should be noted that prepositions meet quite a lot of the criteria listed above. However adverbs, which are often considered as intransitive prepositions, hardly fulfill any of them. We leave it open for the moment whether the assumption that prepositions are defined as major categories should be revised.

[18] A potential exception to this generalization are markers such as English *to* which combine with VPs, i.e., categories which are unsaturated with respect to the subject argument. An even more obvious violation of this rule would be the infinitival marker *zu* in German, which attaches directly to the lexical category. The question is, however, whether the latter should be regarded as a lexical rather than a morphological functional category.

On the other side, comparative markers such as *als* or *wie* appear to fulfill hardly any of these characteristics. In particular, they are quite unrestricted with respect to the categories they combine with, both in terms of their POS and in terms of their saturation. Since the same basically also holds for the closely related coordinative particles, it could be more appropriate to exclude these classes from the above proto-typical generalizations.

To approximate these observations we make the following assumptions.

We analyze functional categories as heads which subcategorize for a substantive category, rather than assuming that the substantive category selects the functional one. Thus, it will be the functional category that requires the presence of a major category while the major category may be unspecified as to whether a functional category has to be present. Verbs, for example, can occur perfectly well without complementizers, whereas complementizers hardly ever go without verbal constituents.[19]

Under the older versions of HPSG, where only one selector feature `SUBCAT` was postulated, one of the arguments for this approach was that we can thus avoid the introduction of an additional type of selection mechanism. Instead, one could have assumed a single homogeneous selection mechanism for all relations between a head and the phrases whose presence is required and licenced by that head. Since the more recent versions already assume a split between `SUBJ` and `COMPS` selections, this homogeneity has been given up to a certain degree.[20] In our description below, we deliberately take one (or two) steps back and operate on the basis of the assumption of a single selector feature for complements. This is mainly to show that the split into the other selector features may be convenient and motivated for other reasons, but that it can be avoided under certain other assumptions.

Nevertheless, functional heads in our view differ from substantive heads in one crucial aspect. Whereas there is hardly any interdependence between the category of a substantive head and the categories it subcategorizes for, we analyze functional heads as operating in a kind of 'parasitic' manner on the major categories that they combine with. Specifically, we posit that *a functional category shares the content of the* `MAJOR` *feature with its complement* and thus adopts the `MAJOR` properties of the substantive category it combines with. The `HEAD` features of a phrase consisting of a functional and a substantive category are thus composed from the `MAJOR` features of the substantive category and the `MINOR` features of the functional category, which crucially can differ from those of its complement.

[19]To capture this phenomenon [Pollard and Sag 1994][37ff] introduce a specific rule schema for head-marker-structures. In our approach, this schema becomes obsolete, since its function can be fulfilled by a simple head-complement-schema.

[20]In fact, the latest version of HPSG, the revisions of "HPSG II" in chapter 9 [Pollard and Sag 1994][362ff], even assumes a third feature `SPR` in parallel to `SUBJ` and `COMPS`.

This assumption has several effects and consequences. Firstly, it will allow us to capture the phenomenon that a phrase headed by a functional head often occurs in the same structural environment as a phrase headed by the corresponding substantive head. However, it also gives us the possibility to constrain functional categories inherently to co-occur only with specific substantive categories. By categorizing a functional category as being of the type *nominal* it is inherently bound to be combined with a nominal substantive category. As we already mentioned above, there is no such comparable regularity observable in substantive categories, which can combine with practically any other category. Nevertheless, since we still assume a split of the HEAD features, this interdependence between MAJOR properties of the substantive and the functional category does not exclude the possibility of the functional head governing properties of the substantive complement. Last but not least, the composition of the HEAD features from two different sources will allow an external governor to subcategorize at the same time for the MINOR properties of the functional head and the MAJOR properties of the major head.

As we will discuss in more detail below, both of these assumptions differ from the treatment of minor categories in [Pollard and Sag 1994][44ff]. The selectional restrictions that we express basically through a head-complement relation are expressed through a specific SPEC Principle in [Pollard and Sag 1994][51]. This principle holds for both determiners, which occur as complements to nouns, and for complementizers, which are not selected at all. Similarly, the fact that functional heads may contribute information to the specification of a maximal projection is captured in [Pollard and Sag 1994] by two different mechanisms, by the *head-marker-schema* in which properties of a MARKER-DTR are percolated to the mother node through the structural specification of the rule, and by the potential that is implicit in the SPEC Principle.

While the above assumptions concern the internal representation of functional categories, we also introduce a concept of *functional completeness* as a primitive. This concept is closely associated with functional categories but concerns primarily the definition of maximal phrasal projections. The basic idea here is that saturation of subcategorization requirements does not always suffice to define a maximal projection, but that the specifier role of functional categories may have to be taken into account. To take a simple example, a finite verbal head in a verb-final clause structure in German may have a fully saturated SUBCAT list, but it will not count as a maximal projection qualifying as a complement to a matrix predicate if it has not first combined with a functional category, such as a complementizer.

Thus, to control those cases where a major category must not occur on its own, as well as the (possible) restriction that a functional category must not occur more than once in a phrase, we introduce a binary valued MINOR attribute FCOMPL, which distinguishes between functionally complete and incomplete categories (123)/(124). In combination with the definition of a maximal projection, the value of this feature will determine whether a lexical or phrasal category still has to combine with a functional head in order to make up a full maximal projection.

(123) $\left[\text{LOC}|\text{CAT}|\text{HEAD}|\text{MINOR}\left[\text{FCOMPL } +\right]\right]_{\textit{f-compl-cat}}$

(124) $\left[\text{LOC}|\text{CAT}|\text{HEAD}|\text{MINOR}\left[\text{FCOMPL } -\right]\right]_{\textit{f-incompl-cat}}$

Maximal projections (126) are then defined as the intersection of functionally complete (123) and saturated categories (125).[21]

(125) $\left[\text{LOC}|\text{CAT}\left[\text{SUBCAT } \langle\ \rangle\right]\right]_{\textit{sat-cat}}$

(126) $\textit{max-cat} \equiv \textit{f-compl-cat} \wedge \textit{sat-cat}$

Analogously, we also define a category *f-inc-sat-cat* which is saturated but functionally incomplete, i.e., which still has to combine with a functional head in order to qualify as a maximal projection.

(127) $\textit{f-inc-sat-cat} \equiv \textit{f-incompl-cat} \wedge \textit{sat-cat}$

In general, we assume that all substantive categories will require their complements to be both saturated and functionally complete. There will be obvious exceptions arising, for example, from the formation of verb complexes in German. Also, we will eventually have to differentiate categories with a saturated COMPS list from fully saturated categories, in order to distinguish between VPs and fully saturated clauses.

For the majority of functional categories we will assume that they are marked in their MINOR attribute as satisfying the requirement of functional completeness. Reversely, functional categories require their complements to be functionally incomplete, which will prevent a minor category from attaching to a phrase which has already been 'closed' by another functional head. However, we also foresee an extension of the notion of functional heads to cover categories such as numerals or specific quantifiers which can co-occur with other functional heads. For these cases, a relaxation of the requirement that all functional heads introduce the property of functional completeness will be necessary.

We summarize the two most crucial assumptions so far in (128) and (129).

[21]For the sake of simplicity we abstract away from the different types of valency saturation with respect to the SUBJ and COMPS attributes.

(128) **Functional Complementation**: In a lexical category of the type *func-cat* the value of the MAJOR attribute is token identical with the MAJOR value of the complement.

(129) **Functional Completeness Constraint**: Every maximal projection is marked as functionally complete in its MINOR feature.

The prototypical definition of a functional category is given in (130) in an abstract type *func-cat*. In this definition we require that functional categories subcategorize for functionally incomplete categories, which contribute the feature MAJOR to the head features of the functional head. For a lexically overt functional category we assume that its single complement has to be fully saturated, i.e., that it has an empty SUBCAT list.

(130)
$$
\textit{func-cat} \;\equiv\; \textit{f-compl-cat} \;\wedge\;
\begin{bmatrix}
\mathtt{LOC|CAT} & \begin{bmatrix}
\mathtt{HEAD} & \begin{bmatrix}\mathtt{MAJOR}\ \boxed{1}\end{bmatrix} \\[1em]
\mathtt{SUBCAT} & \left\langle {}_{\textit{f-inc-sat-cat}}\begin{bmatrix}\mathtt{LOC|CAT}\begin{bmatrix}\mathtt{HEAD}\begin{bmatrix}\mathtt{MAJOR}\ \boxed{1}\end{bmatrix}\end{bmatrix}\end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

As will become clear below, functional completeness is not the only MINOR property that we foresee. Nor do we want to claim that all subclasses of functional categories will be defined for the same set of MINOR features, but rather envisage that different subclasses may be defined for different sets of features, along with features which are common to all functional categories.

4.3.3 Summary

In this section, we outlined our basic assumptions concerning substantive and functional categories. We assumed that categories are basically always defined for two different sets of properties, which we clustered under two HEAD features, MAJOR and MINOR. We specified the main classes of substantive categories and defined their appropriate MAJOR features. We pointed out that the proto-typical functional categories differ from substantive categories in several important aspects, and outlined how these differences can be captured in our approach. This approach is distinct from the traditional HPSG analysis in that it analyzes functional categories as heads which adopt their major properties from their substantive complements. We also introduced a primitive notion of functional completeness, which enters into the definition of maximal phrasal projections. In the next section, we will show these assumptions and definitions can be applied to the domain of nominal constructions.

4.4 Nouns and Determiners

To illustrate the approach outlined in the previous section, we apply the above definitions to the domain of nominal categories. We show how the given inventory of type definitions is employed to categorize the different types of nouns, adjectives and determiners and demonstrate how the different types of phrasal projections fall out from intersecting the abstract types of substantive and functional categories with each other.

4.4.1 Substantive Nominal Categories: Nouns

In the categorizations for nouns (131), we draw on the definition of the type *nominal* (117) and simply add the missing value for `V` to determine the unambiguous part of speech. Adjectives would have the complementary value so that the two types are both defined for the same agreement features, differing only in the value for `V` in their `MAJOR` attributes.

(131) $$noun \quad \equiv \quad nominal \quad \wedge \quad \left[\texttt{LOC}|\texttt{CAT}|\texttt{HEAD}|\texttt{MAJOR}|\texttt{POS} \left[\texttt{V} \; - \right] \right]$$

For simple common nouns, we assume a division into two classes, viz. nouns and noun forms which can occur on their own, such as bare plurals and mass nouns, and singular count nouns, which require the presence of a determiner. According to this distinction, the lexical entries and morphological forms of common nouns are either assigned to a class for which it remains unspecified as to whether the form is functionally complete or not, or to a class which is marked as being intrinsically functionally incomplete.

The former class we assume to comprise the plural forms of count nouns, e.g., *Männer*, *Bücher*, as well as both the singular and plural forms of mass nouns, e.g., *Wasser*, *Wein*, *Weine*. Since these forms can occur equally well with and without a determiner it is left to the structural context to determine the value of the feature `FCOMPL`. They are accordingly assigned the type *noun* (130) with an unspecified `MINOR` feature.

As mentioned before, we envisage that some substantive categories may be intrinsically specified as being functionally incomplete. We make use of this option in the definition of the second subclass of nouns which must combine with a determiner in order to make up a fully saturated maximal phrasal projection. Under this class of intrinsically functionally incomplete categories we subsume all singular forms of count nouns, e.g., *Mann, Buch, Frau.* These forms first have to combine with a functional head or specifier in order to qualify as a phrase which in turn could function as a complement to another substantive category.[22] We mark this property by adding the specification [FCOMPL $-$] to the MINOR feature in the definition of the type *noun*.

(132) *sg-count* $\equiv$ *noun* $\wedge$ *f-incompl-cat*

Analogous to this specification, we define a functionally incomplete saturated category *np*, which has the MAJOR properties of a noun and which serves to specify the type of complements of determiners. We introduce a different type for this category in order to distinguish between syntactically fully saturated, but functionally incomplete categories headed by nouns, and the lexical, potentially unsaturated category of the nouns themselves. The type *np* in (133) can thus be defined as a subtype of the btype *sg-count* intersected with the type *f-inc-sat-cat*, which makes it a fully saturated sign. A schematic instance of the expanded type definition is given in (134).

(133) *np* $\equiv$ *sg-count* $\wedge$ *f-inc-sat-cat*

(134)
$$\left[\texttt{LOC|CAT}\left[\begin{array}{ll}\texttt{HEAD} & \left[\begin{array}{ll}\texttt{MAJOR} & \left[\texttt{POS}\left[\begin{array}{ll}\texttt{N} & +\\ \texttt{V} & -\end{array}\right]\right]\\ \texttt{MINOR} & \left[\texttt{FCOMPL}\ -\right]\end{array}\right]\\ \texttt{SUBCAT} & \langle\ \rangle\end{array}\right]\right]_{np}$$

4.4.2 Functional Nominal Categories: Determiners

For the definition of the functional category 'determiner' and the definition of maximal nominal projections we again draw directly on the type definitions given above. As mentioned, we assume that the functional categories can be classified according to the substantive categories they combine with. Since determiners can be restricted to co-occur with nouns as their complements, the specification of the lexical category of determiners *det* in (135) is thus composed from the type of functional categories, *func-cat* and the type *noun*.

[22]We ignore here cases of type coercion, i.e., cases in which the context enforces a mass noun reading on the count noun:

(i) Es gab Apfel zum Frühstück.

One possible approach to also cover these cases would be to link the count or mass semantics directly to the FCOMPL value of the lexical entry of the singular form. The mass noun reading of a count noun would be chosen only if the noun is marked as functionally complete and *must* not combine with a specifier anymore.

(135) $det \quad \equiv \quad func\text{-}cat \quad \wedge \quad noun$

(136)
$$
{}_{det}\left[\begin{array}{ll} \texttt{LOC|CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \left[\begin{array}{ll} \texttt{MAJOR} & \boxed{1}\left[\begin{array}{ll}\texttt{N} & + \\ \texttt{V} & - \\ \texttt{AGR} & {}_{agr}[\ ]\end{array}\right] \\ \texttt{MINOR} & \left[\texttt{FCOMPL} \ +\right] \end{array}\right] \\ \texttt{SUBCAT} & \left\langle {}_{np}\left[\begin{array}{ll}\texttt{LOC|CAT} & \left[\begin{array}{ll}\texttt{HEAD} & \left[\begin{array}{ll}\texttt{MAJOR} & \boxed{1} \\ \texttt{MINOR} & \left[\texttt{FCOMPL} \ -\right]\end{array}\right] \\ \texttt{SUBCAT} & \langle\ \rangle\end{array}\right]\end{array}\right] \right\rangle \end{array}\right] \end{array}\right]
$$

In the expanded definition of the type *det* (136), the various effects of conjoining these two supertypes are illustrated.

Through the identification of the MAJOR features of the determiner and the noun in combination with the effects of the Head Feature Principle, a phrase headed by a determiner carries the same MAJOR attributes as the noun heading the complement of the determiner. At the same time, we obtain the constraints over the agreement relations between determiners and nouns practically for free, since these properties are part of the MAJOR feature shared between the determiner and its complement.

The information inherited from the abstract type for functional categories automatically restricts the complement of the functional head to be a saturated category which has the MAJOR features of a noun and a MINOR feature which marks it as functionally incomplete; in short, the complement is restricted to a category whose description is met by the type *np*. A projection of the determiner, on the other hand, will be marked as functionally complete.

The subcategorization requirements of determiners according to this specification can thus be met by complements which are either headed by a singular count noun, which matches the selectional restriction [FCOMPL −] in its MINOR attribute, or by a mass or plural noun, which fulfills the requirements by being underspecified in its MINOR attribute. In terms of the type definitions, the latter case follows trivially from the fact that the type *np* is a direct subtype of the type *noun* assigned to mass nouns and plural forms. The former case is guaranteed since there are no incompatible feature descriptions associated with the category *sg-count* and the category *np*. Since the type *np* implies a saturated category, in both cases potential subcategorization requirements of the nouns of course have to be satisfied first.

The restriction imposed on determiners, to combine only with complements headed by a noun, at first sight may appear too strong in the light of constructions such as (137) where a determiner is followed by an apparently verbal construction.

(137) [das [seinen Freunden an Weihnachten eine Freude Machen]]

However, for these constructions a more complex morpho-syntactic analysis seems to be required which also takes into account that the verbal constructions may carry a nominal inflection in the genitive case (138). Unless one wants to accept that nominal inflection can also attach to verbal categories, it does not appear too farfetched to assume that these complements of determiners have been 'transformed' into a nominal category through some morpho-syntactic process.

(138) die Last [des [[allen eine Freude Machen Müssen]s]]

Similarly, it has been claimed in [Haider 1988][42f] that constructions such as (139) and (140) provide evidence that determiners may also take adjectival and prepositional phrases as complements:

(139) [das [mir Anvertraute]]

(140) [der [von gestern]]

This claim appears highly questionable at least for the latter construction, where the determiner (or alternatively an adjective) preceding the prepositional phrase is obligatory, which is neither generally necessary for nominal complements, nor is it even required in the case of (137) and (138). For (139) and in particular for (140) , the assumption of an empty nominal head, as we will suggest below, appears to be more nearly adequate.

4.4.3 Maximal Nominal Projections

The specification of a maximal nominal projection which is both saturated and functionally complete is equally straightforward. We define this type, which we name *dp* for ease of reference, simply as the intersection between maximal projections *max-cat* and the category *noun*.[23]

(141) $dp \quad \equiv \quad max\text{-}cat \quad \wedge \quad noun$

[23] The name of the type *dp* is admittedly somewhat misleading, since the crucial point of this definition is that a subcategorization for a constituent of the type *dp* can be met both by phrases which are headed by a determiner and by determinerless constructions headed by a noun.

(142) $$_{dp}\begin{bmatrix} \texttt{LOC|CAT} \begin{bmatrix} \texttt{HEAD} \begin{bmatrix} \texttt{MAJOR} \begin{bmatrix} \texttt{POS} \begin{bmatrix} \texttt{N} & + \\ \texttt{V} & - \end{bmatrix} \\ \texttt{AGR} \ _{agr}[\ ] \end{bmatrix} \\ \texttt{MINOR} \begin{bmatrix} \texttt{FCOMPL} & + \end{bmatrix} \end{bmatrix} \\ \texttt{SUBCAT} \ \langle\ \rangle \end{bmatrix} \end{bmatrix}$$

Given this final definition we can now draw the map of the various nominal types in a partial lattice (fig. 4.1). This lattice also illustrates that the nominal categories are defined through systematically combining the abstract categorial types with the type *noun*.

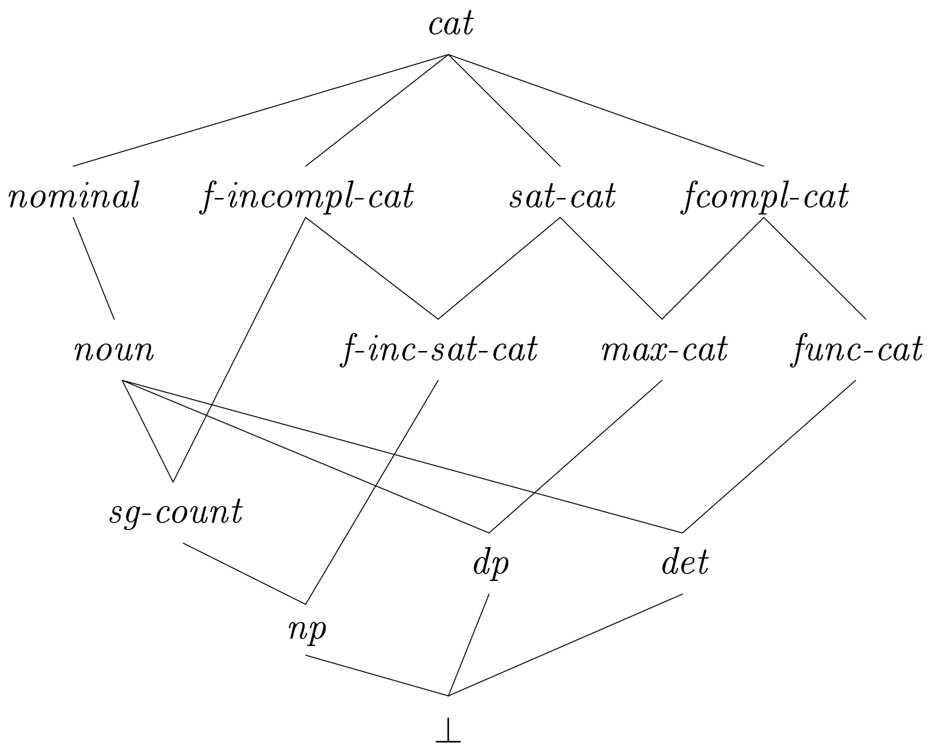
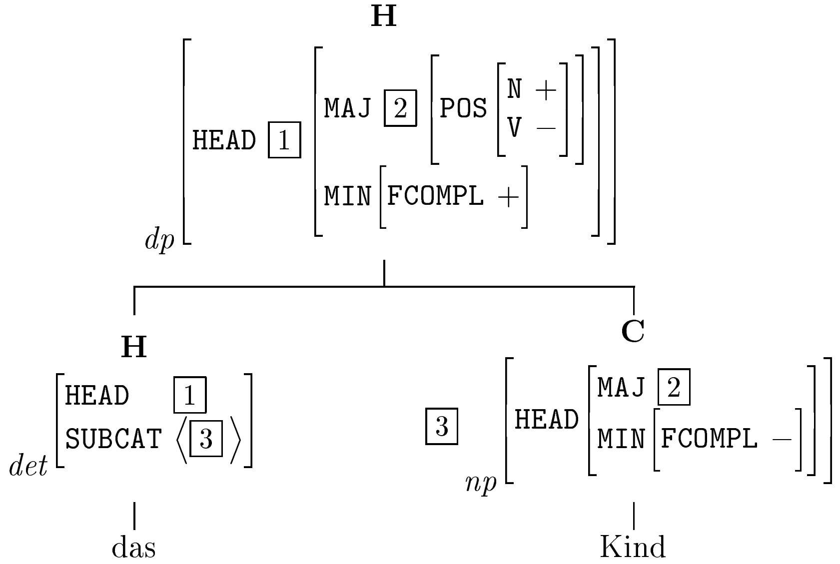


Figure 4.1: Type Lattice for Nominal Categories

As can be easily read off the subsumption relations in this inheritance hierarchy, a maximal nominal projection *dp* according to our definitions need not always be headed by a category of the same sort, but may be projected from either a determiner or a noun.

The first case is illustrated by the structure of a determiner–noun construction (143) which follows from the phrase structure defined by the *head-comp-schema*, the HFP and the SUBCAT Principle. The linear order of the determiner and its nominal complement is captured by the generalization we addressed above, viz. that categories marked [V −] precede their complements. A phrase headed by a determiner qualifies as a maximal projection of type *dp* since its head, the determiner, meets the restrictions on the HEAD features as (i) being functionally complete and (ii) being categorized as a *noun*.

(143)



A determinerless construction headed by a noun fulfills the constraints of a maximal projection *dp* under the condition that the head is not explicitly marked as functionally incomplete. This restriction is trivially met by all phrases which are headed by a category of the type *noun*, which also represents a supertype of the category *dp*, i.e., it includes all (saturated) plural or mass nouns with an underspecified MINOR feature. In contrast, a singular count noun by itself does not satisfy the constraints, since it is marked [FCOMPL −] and therefore has to combine with a determiner first.

Since the lexical entries of plural and mass nouns unify directly with the definition of a *dp*, there is no need for a phrasal projection over the lexical category, if the noun is the only constituent of the phrase (144). To assume a unary phrase structure derivation would be necessary only for an approach which crucially distinguishes between a lexical and a phrasal category. As we showed above in section 3.4, this distinction can be neutralized in our approach and is replaced by other more adequate definitions.

(144)
$$\begin{bmatrix} \texttt{PHON} & \langle \textit{Kinder} \rangle \\ \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CAT} & \begin{bmatrix} \texttt{HEAD} & \begin{bmatrix} \texttt{MAJOR} & \begin{bmatrix} \texttt{POS} & \begin{bmatrix} \texttt{N} & + \\ \texttt{V} & - \end{bmatrix} \end{bmatrix} \\ \texttt{MINOR} & \begin{bmatrix} \texttt{FCOMPL} & + \end{bmatrix} \end{bmatrix} \\ \texttt{SUBCAT} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \\ & dp \wedge noun \end{bmatrix}$$

While the lexical entries of mass and plural nouns match the definition of a maximal projection due to their underspecified MINOR properties, we can extend the lexical classification of nouns to cover also those cases where a noun or nominal category must *not* be combined with a determiner. Examples of such lexical items would be personal pronouns, such as *ich*, *du*, *mir* etc., and possibly also proper nouns. The distribution of these categories is that of fully saturated and functionally complete nominal phrases in most of their usages. We can account for this distribution by assigning the type *dp* directly to pronouns and proper nouns. Such a specification also directly captures the assumption frequently made in DP-approaches that pronouns are nouns and at the same time 'intransitive' or rather fully saturated determiners.

With respect to proper nouns we have to raise the *caveat* that they frequently can also be embedded in phrasal constructions. Even in standard German there is evidence that proper nouns can also be combined with determiners, as for example in *dieser Peter* or *das Frankreich des 20. Jahrhunderts*. We are quite agnostic towards the status of these expressions and simply assume that there may be some proper nouns which never combine with a determiner and that for others there is a strong preference not to.

4.4.4 Summary

The categorial definitions that we specified in this section will correctly account for the distribution of the determiners in the case of bare plurals, mass and count nouns as given in the examples (145) and (146).

(145) der Wein / die Weine / Wein / Weine.

(146) das Auto / die Autos / *Auto / Autos.

We account for this distribution by simply combining the abstract general definitions for substantive and functional categories given in the previous section. Two important properties of our approach are that it extensively exploits the power of underspecification and that it allows to avoid the construction of redundant phrase structure. In comparison to other approaches we can eliminate or avoid a broad range of potentially unnecessary and linguistically little motivated distinctions.

(a) In contrast to structural approaches, there is no need for two different structural rules to derive determined and determinerless constructions.

(b) In contrast to a pure DP-oriented approach where an obligatory determiner position is postulated as providing the structural head of the phrase, we avoid the necessity of empty determiners.

(c) In contrast to an approach which treats constructions with and without a determiner as DPs and NPs respectively and accordingly requires disjunctive subcategorizations on the matrix predicate, maximal nominal projections are uniformly assigned the same type *dp*.

(d) In contrast to an NP-analysis in which the determiner is subcategorized for by the noun, it is not necessary to have a disjunctive subcategorization on the noun in order to express the optionality of the determiner in the case of plurals and mass nouns.

In particular, the comparison with the last approach (d), which forms the basis of the analysis in [Pollard and Sag 1994], illustrates one of the key differences to our analysis. Whereas [Pollard and Sag 1994] choose an approach which is based on disjunction, we chose an approach which is based on underspecification. In this sense our notion of functional completeness, encoded by the `MINOR` feature `FCOMPL`, plays a similar role as the subcategorization for a determiner in their approach. Where the determiner is obligatorily subcategorized for in [Pollard and Sag 1994], nouns are specified as functionally incomplete, i.e., [`FCOMPL` $-$] in our approach. In those cases, where they assume a disjunctive (optional) subcategorization, we assume an unspecified value for the attribute `FCOMPL`. However, this is not the only respect in which our approach may turn out to be a simplification of [Pollard and Sag 1994]. We therefore postpone a more detailed discussion of this analysis, until we have further elaborated our own approach and showed how it will support the analysis of adjectival adjunction, of declination phenomena, and of empty nominal heads.

4.5 Attributive Adjectives and Adjunction

In the previous section, we showed that both determiners and nouns can be treated as heads of nominal phrases, if we give up the assumption that the maximal projection always has to be headed by the same category. The basis for this approach was that maximal projections can be defined independent of the internal structure of a phrase. In this section we investigate how this approach can be accommodated with attributive adjectives as modifiers to nominal heads. The crucial question to be answered is how intermediate projections, or rather the level of projection at which adjectives should adjoin, can be defined in this approach. To explore the possible space of solutions to this problem, we first compare two different approaches to modification in general, viz., the approach taken in [Pollard and Sag 1994] where modifiers are treated as adjuncts and an approach based on assumptions made in categorial grammar, where modifiers are treated as functors or heads.[24]

4.5.1 Modifiers as Adjuncts or Heads

One of the motivations of our analysis of determiners as functional heads is that we can simplify the set of necessary grammatical relations (and thus reduce the set of rule schemata) by treating the relation between specifiers and specified constituents as a head-complement relation. In principle, one could envisage the same for head-modifier relations, which in standard HPSG are treated by a specific head-adjunct-schema, but which one could also attempt to subsume under the general head-complement-relation. In this section, we briefly introduce the standard HPSG approach and explore the problems faced by such an alternative approach, which is very closely modelled on the categorial grammar style treatment of modifiers as functors.

The Head-Adjunct-Schema in HPSG

The standard HPSG approach to adjuncts or modifiers, as it has been outlined in [Pollard and Sag 1994][55ff], is based on the assumption of a separate rule schema for modification, the *head-adjunct-schema*. In this schema, the modified category is analyzed as the head daughter of the phrase providing the syntactic head features while the modifier represents the semantic head of the resulting constituent.

[24]We ignore the meanwhile outdated analysis in [Pollard and Sag 1987] where modifiers are also treated as non-heads but in this case as optional complements of the modified constituent. It can be noted though, that more recently this approach has been taken up in the (e-mail) discussions of the HPSG community again.

The feature structure (147) illustrating this schema consists of a combination of a HEAD-DTR and an ADJ-DTR attribute whose value is taken to be a single sign (rather than a list), which will yield a binary branching structure if this rule is applied recursively. In this schema, the SYNSEM value of the HEAD-DTR is identified with the value of a HEAD feature MOD in the ADJ-DTR which makes the syntactic and semantic information of the HEAD-DTR accessible to the lexical head of the adjunct. As a result, selectional restrictions, such as syntactic agreement or government relations, can be expressed through the lexical entry of the head of the modifier. At the same time, it can be determined lexically how the semantic CONTENT of the HEAD-DTR is to be integrated into the semantic CONTENT of the adjunct.

$$
(147)\quad \left[\begin{array}{ll}
\texttt{SYNSEM} & \left[\texttt{LOC}\left[\begin{array}{ll}\texttt{CAT} & \left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{SUBCAT} & \boxed{2}\end{array}\right]\\ \texttt{CONT} & \boxed{3}\end{array}\right]\right]\\
\texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\texttt{SYNSEM}\ \boxed{4}\left[\texttt{LOC|CAT}\left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{SUBCAT} & \boxed{2}\end{array}\right]\right]\right]\\ \texttt{ADJ-DTR} & \left[\texttt{SYNSEM|LOC}\left[\begin{array}{ll}\texttt{CAT} & \left[\texttt{HEAD}\left[\texttt{MOD}\ \boxed{4}\right]\right]\\ \texttt{CONT} & \boxed{3}\end{array}\right]\right]\end{array}\right]
\end{array}\right]
$$

Note that even in this approach the adjunct fulfills the role of a functor in the sense of a selector. However, this functor or selector role is decoupled from the role of the head. As already mentioned before, this latter role is crucial for determining the head properties of the resulting phrase, which in this case are introduced by the modified constituent.

Modifiers as Heads

An alternative to the above approach would be to treat the adjunct as the head of the phrase both on the semantic and the syntactic level. This approach is basically taken by categorial grammar in which adjuncts are treated as categories of the type X/X, i.e., as functors which take a category X as an argument to yield as a result a category of the type X.[25] The motivation behind this approach

[25]The argument category and the resulting category can be assumed to share most but not all of their properties. [Bouma 1988][39ff] for example suggests that certain modifiers, such as degree adjuncts to adjectives, combine with a lexical category adjective to yield a non-lexical category. Since this latter aspect is equally compatible with the approach of treating adjuncts as modifiers, we ignore this question here.

is mainly that it facilitates the parallel construction of syntactic and semantic functor argument relations.

The main question arising in such an approach is what information can or has to be preserved between the argument and the resulting category. Quite obviously the `HEAD` features of the modifier and those of its complement have to be preserved and should therefore be made coreferent, such that the resulting category will have the same `HEAD` features as the argument of the adjunct by virtue of the HFP. Since adjuncts also must not affect the subcategorization of the complement, the subcategorization requirements of the complement must be preserved by adding them to the `SUBCAT` list of the adjunct. One technique which can be applied for this purpose is the kind of functional composition suggested by [Hinrichs and Nakazawa 1994] for the formation of verb clusters in German, in which the head adds the subcategorization list of its complement to its own subcategorization list. The lexical entry of an adjunctive category functioning as a head then roughly looks as given in (148).

(148)
$$\begin{bmatrix} \texttt{SYNSEM|LOC|CAT} \begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \left\langle \begin{bmatrix} \texttt{LOC|CAT} \begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \end{bmatrix} \end{bmatrix} \bullet \boxed{2} \right\rangle \end{bmatrix} \end{bmatrix}$$

In the context of the structural configuration of a *head-comps-schema* the combination of an adjunct and its argument would have the following representation.

(149)
$$\begin{bmatrix} \texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC} \begin{bmatrix} \texttt{CAT} & \begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \end{bmatrix} \\ \texttt{CONT} & \boxed{3} \end{bmatrix} \end{bmatrix} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{COMP-DTR} & \begin{bmatrix} \texttt{SYNSEM} \; \boxed{4} \begin{bmatrix} \texttt{LOC|CAT} \begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \texttt{HEAD-DTR} & \begin{bmatrix} \texttt{SYNSEM|LOC} \begin{bmatrix} \texttt{CAT} & \begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \left\langle \boxed{4} \bullet \boxed{2} \right\rangle \end{bmatrix} \\ \texttt{CONT} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

However, there are several questions arising from this approach which do not have trivial and straightforward answers. The first problem concerns the adequate distinction between ‘genuine’ complements of adjuncts and the complement which represents the modified argument. Instances where heads of modifier phrases subcategorize for other complements themselves are standardly prepositions, which first have to combine with a noun phrase to produce a phrasal adjunct, but also certain adjectives which can combine with dative or genitive objects (e.g., *treu*, *überdrüssig*). Therefore, the schematic representation of the lexical entry of the modifier head in (147) appears to be too simple, and has to be refined such that the SUBCAT list of the argument is appended to the list of the adjunct's 'genuine' subcategorization. The structure of a modifier head would then roughly be as given in (150), where the dots represent the adjunct's 'genuine' complements.

(150) $$\left[\texttt{SYNSEM|LOC|CAT}\left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{SUBCAT} & \left\langle \ldots, \left[\texttt{LOC|CAT}\left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{SUBCAT} & \boxed{2}\end{array}\right]\right] \bullet \boxed{2} \right\rangle\end{array}\right]\right]$$

However, this does not solve the problem entirely, since one still has to guarantee that the modifier's genuine complements are consumed before the modifier combines with the modified argument. Such a constraint would basically require a restriction on the subcategorization principle to the effect that cancellation of arguments always concerns the first element(s) of a SUBCAT list. However, such an assumption would be incompatible with the approach in which the complements can be taken off the SUBCAT list in any order for free word order languages. Although there may be some languages, as for example Latin, which do in fact allow a freedom of word order to the extent that complements of modifiers and complements of modified heads are arbitrarily permuted, this is typically not the case for languages where scrambling is constrained to complements within one head domain. Since the head domain of the modifier head is extended in the above representation to include also complements of the modified head, the possible solutions to the problem are quite limited.

One possible approach was suggested to us by Hans Uszkoreit and is based on the assumption that the scrambling effects are obtained by a lexical rather than by a structural operation on the subcategorization list. A 'structural' solution to free word order is based on the assumption that *any* element from the subcategorization list can be cancelled in a head complement configuration. In a lexical approach to scrambling the possible permutations of the complements are precomputed in the lexicon, such that cancellation can always affect the first element on a SUBCAT list. It it not quite clear, however, whether this approach could introduce some unforeseen (practical rather than theoretical) complexity into the processing of free word order.

Another alternative would be to separate the two types of subcategorization requirements, such that one attribute captures the genuine subcategorization of the modifier and a second additional feature, say MOD-COMP, expresses the selection of the modified argument conjoined with its complement. This means that one has to introduce an additional instance of a very specific and rather exceptional *head-complement schema* which combines a `HEAD-DTR` with an empty `SUBCAT` list with a `COMP-DTR`, which is selected through the `MOD-COMP` feature. Not only would this schema be extremely similar to the *head-adjunct schema*, it would also require that the standard Subcategorization Principle is locally disabled for this particular rule schema. Since the modifier has to preserve the subcategorization requirements of the modified head, the `SUBCAT` list of the `COMP-DTR` has to be passed on to the mother, i.e., the empty `SUBCAT` list of the head daughter would be turned into a non-empty one on the mother node, as illustrated in (151).

(151)
$$\left[\begin{array}{ll}
\texttt{SYNSEM} & \left[\texttt{LOC} \left[\begin{array}{ll} \texttt{CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \\ \texttt{MOD-COMP} & \langle\ \rangle \end{array}\right] \\ \texttt{CONT} & \boxed{3} \end{array}\right]\right] \\
\texttt{DTRS} & \left[\begin{array}{ll} \texttt{COMP-DTR} & \left[\texttt{SYNSEM}\ \boxed{4} \left[\texttt{LOC|CAT} \left[\begin{array}{ll} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \end{array}\right]\right]\right] \\ \texttt{HEAD-DTR} & \left[\texttt{SYNSEM|LOC} \left[\begin{array}{ll} \texttt{CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \langle\ \rangle \\ \texttt{MOD-COMP} & \left\langle \boxed{4} \right\rangle \end{array}\right] \\ \texttt{CONT} & \boxed{3} \end{array}\right]\right] \end{array}\right]
\end{array}\right]$$

In short we would end up with no gain in terms reduction of phrase structure schemata, but would have lost quite some elegance and simplicity in terms of general principles.

The second problem of an analysis of modifiers as heads is related to word order. In the head-adjunct approach certain word order generalizations concerning the position of the head relative to its arguments or adjuncts can be expressed by referring to the head features of the head. These generalizations cannot not be maintained, if modifiers as functors are head daughters and have the same head features as their arguments, since for these configurations the predictions of the original principle are reversed. For example, if one wanted to express the generalization that a verbal head appears to the right of its complements and (modifying) adjuncts in German, this would have to be specified such that a head daughter with verbal head features appears to the right, if it always was a verb, but not if it has obtained the head features from a verb due to its grammatical function as an adjunct—a constraint which does not come out very naturally in a unification-based framework.

The ultimate reason for this problem is already a problem by itself. Above, we were assuming that categories can be characterized in terms of a part-of-speech feature, which can be based on morpho-syntactic properties and which is above all independent of the possible grammatical functions of a category. In a categorial grammar style approach such an independent classification is impossible if the heads of adjuncts are defined in terms of the `HEAD` features of their arguments—an assumption which is quite essential since the category of a modified constituent has to be preserved. In our approach, we could define linear precedence as a function of the part-of-speech of a category and its grammatical function. For example, the statement that adjectival heads typically occur in a right-peripheral position refers to both the part-of-speech 'adjective' and the grammatical function 'head'. The same could hold for other LP statements ordering the sequence of complements, e.g., that NP complements precede PP complements. However, such LP statements are considerably more difficult to express in an approach in which only the grammatical function but not the category are specified explicitly.

We finally only briefly mention a third problem, without going into further details. If modifiers are analyzed as heads, the analysis of non-local dependencies becomes more complicated if modifiers are involved. If a modifier itself is dislocated, this construction will have to be treated as an instance of *head-movement* with all its potential consequences and problems. On the other hand, if some constituent functioning as an argument to a modified head is dislocated, this could mean that there has been an extraction out of an modifier's argument position. This will complicate or at least weaken the generalizations on long-distance extractions, which may be desirable again for some few cases, such as preposition stranding in English, but which is clearly in contrast to the evidence in German, where we can extract out of complement clauses, but not out of adverbial clauses.[26]

Thus, since many of these problems appear to be too serious to compensate for the potential advantages of analyzing modifiers as heads, we will maintain the standard HPSG analysis of adjuncts as non-heads. However, we will refer to this alternative approach, when it can provide solutions to problems which could be superior to the standard approach.

4.5.2 Adjectival Adjuncts and 'Submaximal' Projections

Following the standard HPSG approach to adjunction, we are above all faced with the question how the selectional constraints introduced through the `MOD` feature of an attributive adjective should be formulated. As it turns out, this question raises a non-trivial problem not only for our approach, but also for the analysis of adjectives sketched in [Pollard and Sag 1994] and for the categorial grammar

[26] For a description of the problem and an LFG-based analysis cf. [Netter 1988].

approach by [Bouma 1988]. The central dilemma arises from the requirement that the adjunct must be able to select the appropriate projection level of a nominal head, while it must not introduce any restrictions on this head which could have undesirable side effects beyond the local attachment domain. Specifically, we have to make sure that the underspecification of the nominal substantive heads is not overwritten and overspecified by any constraints imposed on the modified head by the adjective's selectional feature.

Overgeneration vs. Overspecification

Selectional constraints imposed on the modified constituent in a head-adjunct configuration will normally be rooted in the lexical type of the adjunct's head, in our case the attributive adjective. As a minimal, however insufficient, constraint, we can postulate that an adjectival adjunct will select a modified category with the MAJOR features of a *noun*.

(152)
$$
{}_{adj}\begin{bmatrix} \texttt{LOC|CAT|HEAD} & \begin{bmatrix} \texttt{MAJOR} & \begin{bmatrix} \texttt{POS} & \begin{bmatrix} \texttt{N} & + \\ \texttt{V} & + \end{bmatrix} \end{bmatrix} \\ \texttt{MOD} & {}_{noun}\begin{bmatrix} \texttt{LOC|CAT|HEAD|MAJOR} & \begin{bmatrix} \texttt{POS} & \begin{bmatrix} \texttt{N} & + \\ \texttt{V} & - \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

As it stands, this specification will lead to overgeneration, since an adjective of this type can combine with any constituent which is lexically headed by a category with the major head features of a noun. Such a constituent could correspond to the type *np* (153), or to a fully saturated maximal projection of type *dp*, which may or may not contain a determiner (154)/(155).

(153) der gute Wein

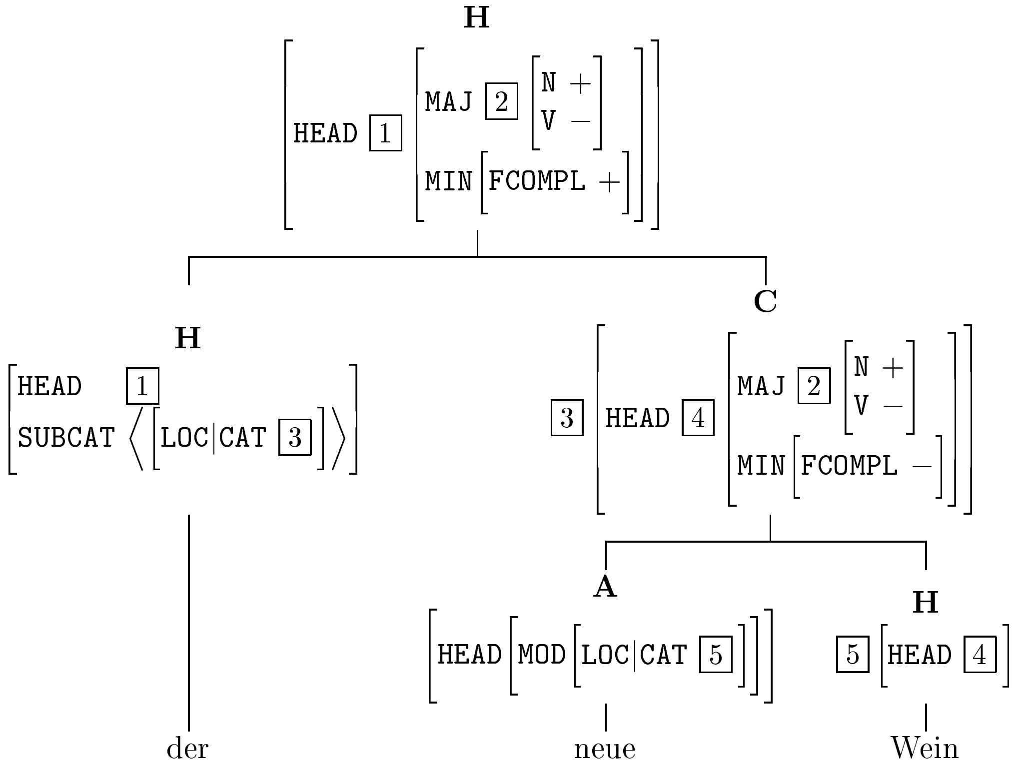
(154) guter Wein / gute Weine

(155) *gute der Wein / *guter der Wein

Whereas coverage of the first two examples (153)/(154) is clearly desirable, the exclusion of the ungrammatical construction (155) appears to be impossible, at least on the basis of the distinctions introduced so far. We will discuss these cases in turn in order to illustrate the problem in more detail.

Example (153) illustrates the case where the adjective adjoins to a nominal head which is marked as functionally incomplete. This configuration is represented by (156), where the adjective combines with a noun to form a category of type *np* which then attains functional completeness through combination with the determiner. Note that the marking of the *np* as functionally incomplete in this case neither derives from the noun itself, which—being a mass noun—was underspecified in the lexicon, nor from the adjective's selection feature, but from the determiner.

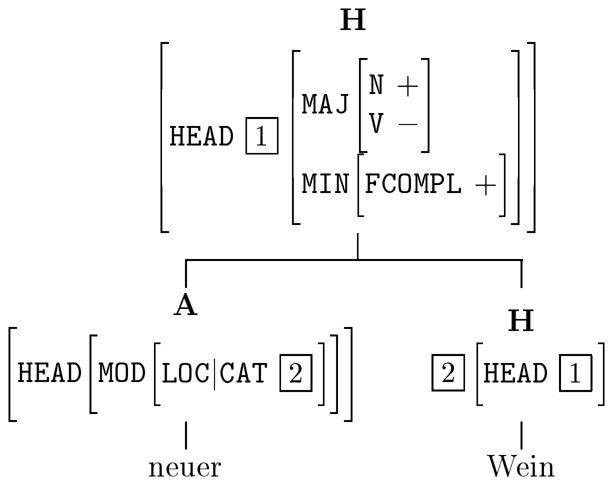
(156)



The second example stands for the well-formed configuration where an adjective adjoins to a category which is marked as functionally complete, as given in (157). In this case, the mass noun is the head of the entire maximal projection. Again it is underspecified with respect to functional completeness in the lexicon, but receives this marking through the type definition of the maximal projection *dp* and the percolation of the HEAD features.

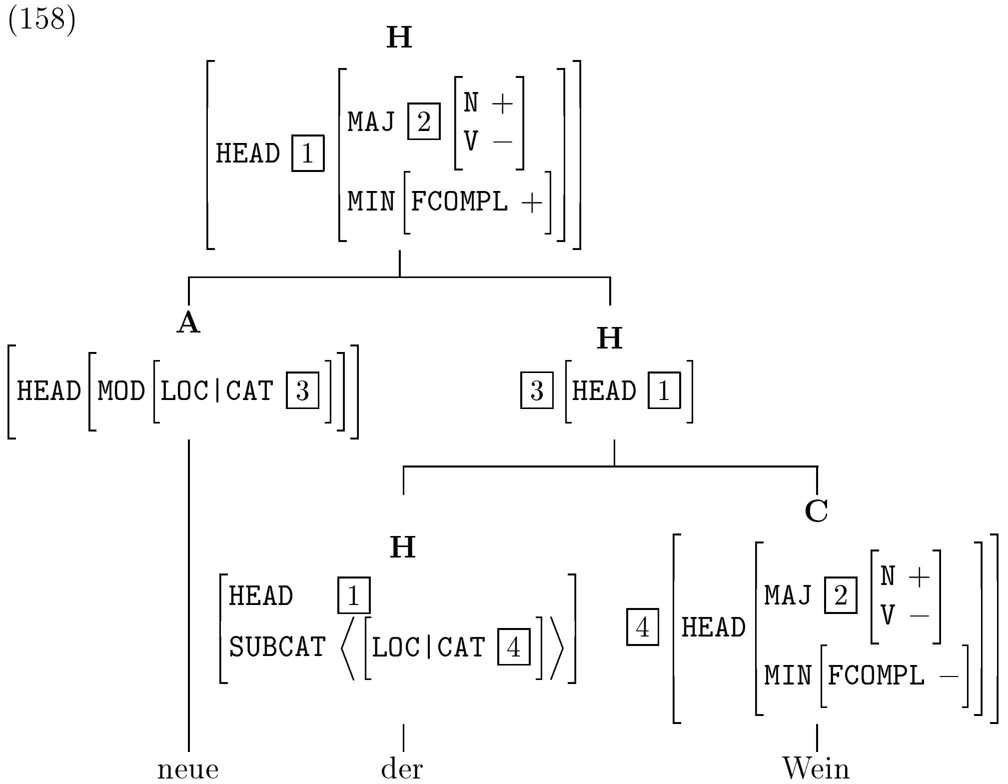
The crucial dilemma becomes clear in the comparison of the grammatical structure (157) with the ungrammatical (158). In both cases, the MINOR feature of the constituent which is selected by the adjective is marked as functionally complete. However, it is impossible to distinguish the source of the specification [FCOMPL +] locally at the adjunction site of the modifier, since this feature percolates along the projection of the head according to the HFP. Thus, the minor property of functional completeness can have been introduced by the type definition *dp* of

(157)



the maximal projection into which the head noun eventually projects in an altogether determinerless construction (157), or it can be introduced by a determiner dominated by the category to which the adjective attaches (158).

(158)



To block derivations like (158), it may come to mind to constrain the MOD value of the adjective such that it can only modify a constituent of type *np*, i.e., that it can only select a determinerless constituent which is marked as functionally incomplete.

However, in our analysis, such a constraint would inadvertently also lead to undergeneration, as it would block the derivation of determinerless constructions with adjectival modifiers as maximal nominal projections. Since the adjective will mark the head of the noun phrase invariably as functionally incomplete no matter whether it was lexically specified as such or not, it can overspecify the noun's minor features and thus cause an undesirable side effect. Any projection of an originally underspecified mass or plural count noun would be marked as functionally incomplete as soon as it has combined with an adjective.

If we pursued such an approach and wanted to compensate for this overspecification, we would be ultimately forced to revert to some structurally oriented analysis, in which the maximal projection of a determinerless construction is derived either through a unary structural rule or the introduction of an empty determiner. Under such assumptions, the property of functional completeness would be almost reduced to a feature encoding the bar-level of projections as we know it from $\overline{\mathrm{X}}$-syntax. Of course, even if we opted for such a solution, we could still claim to have gained some insight into the categorization of functional heads through the distinction of `MAJOR` and `MINOR` head features. However, we would also abandon the theoretical rigour we derive from forbidding unary rules for the mere purpose of category transformation and from eliminating syntactically unmotivated empty terminals.

Functionally Complete vs. Specified

To avoid such a 'structural solution', we take a different tack and reconsider the status of the property of functional completeness. In the light of the problem we just described, it becomes clear that this property is relevant for the 'external' definition of maximal projections, but that it should be considered as logically independent from the properties which play a role in the 'internal' structure of a phrase.

We introduced the notion of functional completeness to encode whether a noun may function as the head of an *np*, of a *dp*, or whether it is unspecified with respect to this distinction and may function as the head of either type. In this sense, functional completeness describes whether or not a constituent still has to combine with a functional head to qualify as a maximal projection. Clearly, if we treat this property as a lexically determined `HEAD` feature and if we want to exploit the effects of underspecification, it must not be affected (and specified) by adjunction but remain invariant within a head domain. As a consequence, the phrasal configuration to which an adjective may attach cannot be defined through functional completeness, since both determined and undetermined constituents may be marked as functionally complete, while adjectival adjunction has to be sensitive to whether or not a determiner occurred in the modified constituent.

Although functional completeness fails to provide the necessary distinction, we can still exploit the fact that determiners cause a partial discontinuity in the percolation of head features from the substantive nominal head. Thus, to encode the distinction between specified and unspecified constructions, we can introduce a second binary-valued `MINOR` feature [`SPEC` $+/-$], short for `SPECIFIED`.[27]

This feature *must* be a `HEAD` feature, if we want to introduce it lexically, since only `HEAD` features percolate from a lexical head daughter to the mother node. If `SPEC` were not a `HEAD` feature or if the determiner did not function as the head in a nominal phrase, either a specific rule schema or a specific structural principle would be required to achieve the appropriate result. Furthermore, it is crucial that we can draw on a division of `HEAD` features and that `SPEC` is a `MINOR` feature, since categories headed by a functional head differ from categories headed by a substantive category only in this respect.[28]

On the basis of these assumptions, the necessary modifications of the previously given specifications can be limited to the definition of lexical types, affecting only the category definitions for determiners and adjectives. In the case of determiners the feature [`SPEC` $+$] indicates the presence of a specifier (159).

(159)
$$ {}_{det}\left[\begin{array}{ll} \mathtt{LOC|CAT} & \left[\begin{array}{ll} \mathtt{HEAD} & \left[\begin{array}{ll} \mathtt{MAJOR} & \boxed{1}\left[\mathtt{POS}\left[\begin{array}{ll}\mathtt{N} & + \\ \mathtt{V} & -\end{array}\right]\right] \\ \mathtt{MINOR} & \left[\begin{array}{ll}\mathtt{FCOMPL} & + \\ \mathtt{SPEC} & +\end{array}\right] \end{array}\right] \\ \mathtt{SUBCAT} & \left\langle {}_{np}\left[\mathtt{LOC|CAT|HEAD}\left[\begin{array}{ll}\mathtt{MAJOR} & \boxed{1} \\ \mathtt{MINOR} & \left[\mathtt{FCOMPL}\ \ -\right]\end{array}\right]\right]\right\rangle \end{array}\right] \end{array}\right] $$

In the definition of adjectives (160), the converse constraint is added to the selectional restrictions imposed by the `MOD` feature on the nominal head, such that this head has to be compatible with the specification [`SPEC` $-$], which is not satisfied by a constituent headed by a determiner.

[27]This feature should not be confused with the complex valued `SPEC` feature introduced by [Pollard and Sag 1994] on head nouns to express the subcategorization for a specifier.

[28]As we will show below, such a 'lexical' solution is not available to the approach taken by [Pollard and Sag 1994]; it would require some changes affecting the structural context of specifiers.

(160)
$$
{}_{adj}\left[\texttt{LOC|CAT|HEAD}\begin{bmatrix}\texttt{MAJOR}\begin{bmatrix}\texttt{POS}\begin{bmatrix}\texttt{N} & + \\ \texttt{V} & +\end{bmatrix}\end{bmatrix} \\ \texttt{MOD}\ {}_{noun}\left[\texttt{LOC|CAT|HEAD}\begin{bmatrix}\texttt{MAJOR}\begin{bmatrix}\texttt{POS}\begin{bmatrix}\texttt{N} & + \\ \texttt{V} & -\end{bmatrix}\end{bmatrix} \\ \texttt{MINOR}\begin{bmatrix}\texttt{SPEC} & -\end{bmatrix}\end{bmatrix}\right]\end{bmatrix}\right]
$$

All phrase structure schemata and principles, as well as the other type definitions given above remain unaffected altogether.

In the light of the alternative structural approach sketched above, one might argue that the suggested solution comes down to pure 'feature hacking'. In particular, one might raise the objection that two bits of information (the values of the attributes FCOMPL and SPEC) are needed in this approach to encode a property which on the surface appears to be equally well expressed by a 'simple' structural bar–level. However, it has to be kept in mind that in a structural approach one also needs two bits of information for the derivation of determinerless constructions, namely the $\overline{\mathrm{X}}$-level (corresponding to our SPEC feature) *plus* a unary structural rule or an empty determiner, which allows to raise the phrase to a maximal projection (corresponding to our notion of functional completeness).[29] In fact from this point of view, it can be seen as a drawback of the structural approach that the two bits of information, which we argued to encode externally and internally relevant properties, cannot be separated from each other in an $\overline{\mathrm{X}}$-oriented approach, but that they always have to be fully specified and that they enforce a particular analysis for determinerless constructions.

Thus, the difference between the two approaches lies not so much in the amount of information to be specified, but rather in the distinction between the same information being expressed *structurally* rather than *lexically*, as well as in the choice of employing *disjunction* rather than *underspecification* to capture the relevant generalizations. Since HPSG is designed as a lexicalist constraint-based framework, we can claim that the respective latter options are more nearly adequate, since they are constrained to a more local (lexical) domain and at the same time are less redundant.

It should be mentioned, however, that we use the distinction between 'lexical' and 'structural' approaches not in the sense that certain information should preferably

[29] We leave it to the reader to verify that the often suggested encoding of $\overline{\mathrm{X}}$-projections by means of two features [MAX +/−] and [MIN +/−] will not solve the problem either. This sort of encoding at some point has been quite common in the G&B paradigm. Cf. among others [Kolb and Thiersch 1991][220ff].

always reside in the lexicon, but rather in the sense that no additional structural device is introduced. Lexicalization of information is only preferable over the postulation of some specific rule, but this does not mean that true generalizations should not be located in a structural rule where this is possible. Thus, we would of course consider it as an advantage, if the negative `SPEC` feature occuring in the lexical category of the adjectival adjunct could be generalized to all other adjuncts. In this case, a more appropriate location of that information would be the structural schema for head-adjunct configurations. We will further investigate extensions of this sort below, when we come to complementizers as a related class of functional heads.

Note, finally, that the problem of adjunction we discussed here is by no means bound to our specific approach to functional heads. As we will show in more detail in section 4.9, exactly the same dilemma arises if we opt for an NP-analysis and follow the suggestion by [Pollard and Sag 1994] to treat determiners as subcategorized complements of the noun.

4.5.3 Summary

In this section, we discussed the issue of modification and adjunction in more detail. We first explored the possibilities to reduce the set of grammatical relations and phrase structure schemata further by substituting the *head-adjunct-schema* by an analysis in which modifiers take the role of functors and the modified constituents the role of an argument. Such an analysis, which would be closely modelled on the categorial grammar approach, would mainly make sense if the functor role could be identified with the function of a head-daughter. As it turns out this neutralization of the distinction between *head* and *functor* or *selector* would cause a number of serious problems, which make such an analysis appear considerably more clumsy than the standard HPSG approach.

We then returned to our analysis of functional categories and discussed the question of how the selectional restrictions of an adjoining attributive adjective have to be specified to avoid overgeneration of a certain kind. We showed that the distinctions introduced so far did not allow to maintain an analysis of determinerless constructions on the basis of underspecification and at the same time to prevent adjunction to maximal projections. Rather than resorting to a structural solution with unary rules or empty determiners, we exploited the split of head features and introduced a second `MINOR` feature `SPEC` which encodes the distinction between specified and unspecified constituents. We argued that functional completeness, as an externally relevant property, should be seen as independent of this phrase-internally relevant feature. Also, since this solution can be constrained to the lexicon and since it supports a strategy of underspecification, we consider this approach to be more in line with the basic philosophy of HPSG.

4.6 Empty Nominal Heads

In this section, we extend our fragment of nominal phrases to cover constructions with a missing head noun. We first present some of the relevant data which in our view motivate the postulation of an empty nominal head and discuss the problems arising from these constructions. We then present two alternative approaches, based on two different generalizations. The first solution operates on the assumption that the *left periphery* of a constituent of the type *dp* must not be empty, the second one is based on the generalization that the *head* of a *dp* must not be empty. The second approach will presuppose the alternative analysis of attributive adjectives functioning as heads in nominal constructions, described in the previous section. We finally show how the distribution of some specific forms of inflected and uninflected determiners can be described and integrated into our approach.

4.6.1 Motivation for Empty Nominal Heads

In the following, we discuss some of the relevant data which involve nominal phrases with the head noun missing. Since we take the point of view that empty nodes should not be introduced without substantial motivation, the purpose of this discussion is to investigate and evaluate at least some of the most salient alternative approaches.

Some Basic Data

Constructions with an empty nominal head in German are illustrated by the following simple examples, which only accidentally involve coordination which is not a prerequisite for the occurence of the phenomenon.

(161) das rote Auto und das blaue **e**

(162) rote Autos und blaue **e**

(163) gesalzene Butter und ungesalzene **e**

These constructions very closely correspond to English nominal constructions with so-called '*one* anaphora', as illustrated in (164)/(165). The only difference appears to be that *one* anaphora are ungrammatical in mass noun constructions in English (166) and that *one* anaphora can occur on their own without any modifiers or specifiers (167).

(164) the red car and the blue *one*

(165) red cars and blue *ones*

(166) I like salted butter, my wife prefers unsalted **one*

(167) He asked me for a book and I gave him *one*

Very similar to English, there are very few restrictions on the context in which a head noun can be eliminated in a nominal construction in German. Basically a nominal head can be left out in any construction where a modifier or specifier precedes the noun.

(168)	der kleine Mann	vs.	der kleine **e**
	kleine Männer	vs.	kleine **e**
	diese Männer	vs.	diese **e**
	die beiden Männer	vs.	die beiden **e**
	beide Männer	vs.	beide **e**

Same as for *one* anaphora in English it is not necessarily only the noun that can be interpreted as missing, but other (immediately preceding) material can be involved as well. Yet, even in close coordination there is hardly any evidence that what has been eliminated must correspond directly to a substring in the preceding conjunct (170).

(169) der kleine alte Mann und der große **e**

(170) jene kleinen alten Männer und dieser **e**

One of the most essential restrictions that has to be accounted for is that the empty noun must be preceded by some constituent in the nominal phrase, be it determiner or adjective, as it is given by the contrast between (171) and (172).

(171) [Alte Männer mit Hut] haben [junge **e** mit Mütze] getroffen.

(172) [Männer mit Hut] haben * [**e** mit Mütze] getroffen.

In addition, it appears to be a necessary condition that some of the preceding material carries a morphologically marked nominal inflection, which can be illustrated by a minimal pair involving the inflected (173) and uninflected forms (174) of the quantifier *alle*.

(173) [alle diese Männer mit Hut] und auch [alle **e** mit Mütze]

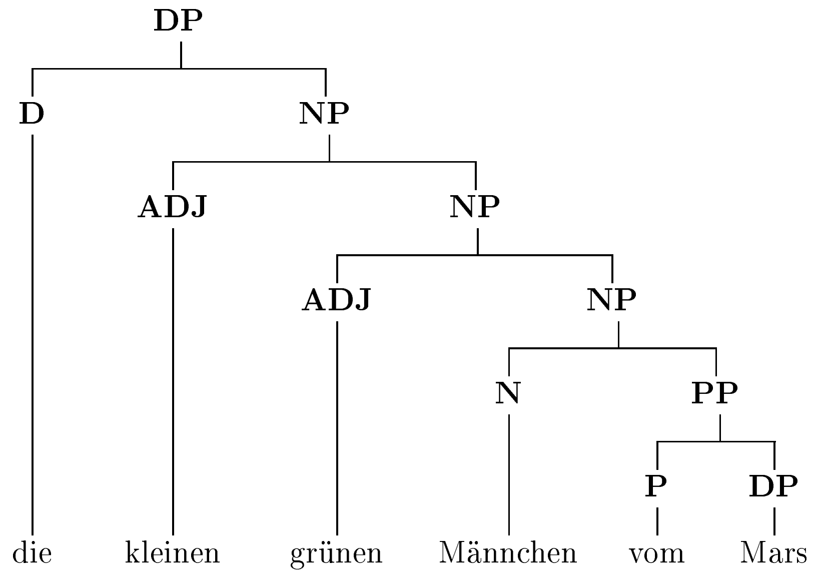
(174) [all diese Männer mit Hut] und auch * [all **e** mit Mütze]

One of the most crucial questions in the analysis of these constructions, which has to be answered first, is whether one should assume the presence of an empty nominal head, analogous to the English *one* anaphor, or whether some other approach would be more adequate. We will discuss two alternative solutions, namely the assumption of some *nominalization* process and the assumption of some *coordination ellipsis*.

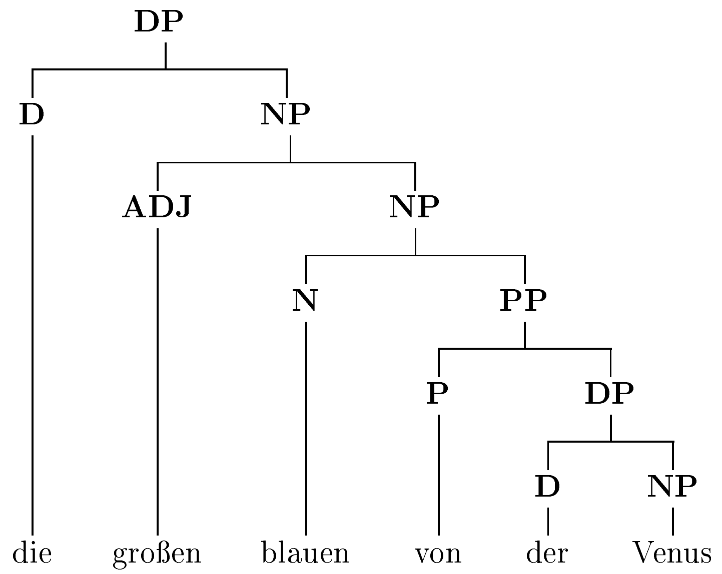
Nominalization

The first alternative to an empty head noun is based on the assumption that the last constituent, be it determiner, adjective or numeral, preceding the missing noun underwent some process of nominalization and functions as the head noun of the phrase. Some illustrating structures are given in (175) and (176).

(175)



(176)



This approach has the advantage that every nominal phrase has to contain a substantive nominal head. Such a constraint would solve the problem of non-existent pre-nominal determiners or adjectives, since constructions consisting only of a post-nominal modifier would be marked as ungrammatical for not having a substantive nominal head.

Technically, such an approach would work comparatively straightforwardly for the simple cases, if one assumed some lexical rule which took a determiner, adjective or numeral as input, and produced as output a noun removing the respective `SUBCAT` or `MOD` feature from the original category in the process. However, this lexical rule would also have to cover those constructions involving a complex adjectival phrase or participial construction as in (177) and (178).

(177) [die [[ziemlich alten] Männer] und [die [besonders jungen]]

(178) [die [[ständig Pfeife rauchenden] Männer] und
[die [sich darüber ärgernden]]

The closest solution for these constructions would be again the postulation of a unary rule which performed the respective category transformation. As we already pointed out in our introductory section 4.2 such unary rules are not admissible in HPSG, if a monotonic Head Feature Principle is to be maintained.

A second set of examples which could be a problem for such a nominalization hypothesis are cases of nouns with a specific syntactic subcategorization frame which is inherited by the empty head.

(179) [Die [starke [Konzentration [auf die Wirtschaft]]]] und
[die [weniger große [auf den Umweltschutz]]]

(180) [Das [erste [Zug [nach München]]]] fährt um sechs Uhr,
[der [letzte [nach Berlin]]] um Mitternacht.

Since these complements do not exhibit a different behaviour in an elliptical and non-elliptical case, the general phrase structure set-up will require some kind of nominal head to be present. Under the nominalization hypothesis one has to stipulate that the respective determiner, adjective, etc. not only adopts the precise function (and category) of the noun, but also inherits the subcategorization frame of the antecedent. Clearly, whether or not this will be a problem will depend on how complements of nouns are treated, i.e. whether one assumes a standard subcategorization mechanism or rather some mechanism based on semantic frames.

A somewhat related problem is that properties, such as the contrast between mass and count nouns must be inherited by the construction with the missing noun. Quite obviously, examples such as (181) are ungrammatical if the second, anaphoric construction is missing a determiner in the singular.

(181) Ich wollte mir ein blaues Auto kaufen,
Hans dachte an *grünes / ein grünes / grüne.

Again, the nominalized head of the second construction must be constrained in some way to provide the relevant properties.

These constructions illustrate one of the main deficiencies of the nominalization approach, viz., that it fails to capture in a straightforward way the obvious anaphoric relationships involved in these constructions. Rather than limiting the set of anaphora to a small closed class of elements (lexically realized or not), this approach implies that there is an open set of anaphora which could come in the form of practically any adjective or participle constructions that happens to occur last in an elliptical nominal construction.

For these reasons we conclude that, at least for the complex cases involving adjectives, a nominalization approach does not provide an obviously adequate analysis. It may require the postulation of unary rules again, which will be inconsistent with the theory, and the assumption of a lexical rule will not require the postulation of an open class of anaphora. Of course, we have to admit that the proposed alternative of an empty head noun will not solve the problem of anaphoric relations either. However, as we pointed out above in section 4.2, it will at least provide a theory consistent solution and it may help to constrain this latter problem to a more local domain.

We leave it open for the moment, whether such a nominalization approach will be feasible for the constructions involving only determiners or specifiers. This possibility will be discussed below.

Coordination Reduction

Another alternative approach is based on the assumption that the ellipsis of the nominal head can be reduced to some type of coordination reduction. Most of the ellipsis phenomena occurring in coordination in German can be reduced to three basic constructions and their combinations.[30] These basic constructions are

(a) Left Periphery Deletion: In these constructions material at the left periphery of the second conjunct is deleted. Under certain assumptions, these constructions can also be reduced to simple constituent coordination.

(182) Diesen Wagen hat sich [Hans gekauft] und [Peter ausgeliehen].

(b) Right Periphery Deletion: These constructions involve ellipsis of an arbitrary string at the right periphery of the first conjunct. The phenomena appear to be very closely related to Right Node Raising, which is unlikely to provide an adequate analysis for German, however.

(183) Hans wird sich ein zwei- und
Peter will sich ein viertüriges Auto kaufen

[30] Cf. [Höhle 1991], who argues that only two of these should be regarded as genuine elliptical constructions.

(c) Gapping: While in the first to types of ellipsis coherent strings are affected, gapping has to be formulated on the basis of grammatical dependencies. Crucially it involves the ellipsis of a finite verb together with some of its complements or adjuncts. In addition directly dependent non-finite verbs together with a subset of their complements can be deleted.

(184) Peter will die Maria überreden, dem Hans ein Auto zu verkaufen und Fritz den Paul dem Emil.

Constructions involving a missing head noun do not follow either of these types of ellipsis in the way they are described. They could have in common with the Left and Right Periphery Deletion that in most cases some coherent material is elided. However, there are also constructions where discontinuous substrings could be reconstructed.

(185) Peter hat sich [einen *viertürigen* blauen *Wagen*] gekauft und
Paul [einen roten].

The elliptical construction in (185) undoubtedly also has an interpretation as *einen viertürigen roten Wagen*. In addition, nominal ellipsis differs from the Periphery Deletions in that the deleted material must potentially form a coherent phrase in terms of grammatical dependencies. In this sense, nominal ellipsis is much closer to gapping, which also involves a head plus some constituents dependent on the head.

It is interesting to note in this context that genuine verbal gapping can also interfere with the constraint mentioned above, viz., that constructions with an empty nominal head must not have an empty left periphery. Apparently, elision of the head noun together with all pre-modifiers is considerably more acceptable if this elision goes together with a verbal gapping constructions (186).

(186) Hans hat [die Kekse [von seiner Mutter]] aufgehoben,
Peter [**e** [von seiner Großmutter]].

In analogy to the gapping constructions where a complete complement of an embedded verb is deleted, the gapping has to involve the finite verb and all verbs on which the noun phrase is dependent, be it directly, or indirectly via a chain of government dependencies. If only one verb (187) or even no verb (188) is elided the construction becomes ungrammatical again.

(187) Hans hat die Kekse von seiner Mutter aufgehoben,
*Peter von seiner Großmutter aufgefuttert.

(188) Hans hat die guten Kekse von seiner Mutter aufgehoben,
*Peter hat von seiner Großmutter aufgefuttert.

Nevertheless, there is one crucial difference between the coordination ellipsis and the missing head nouns which concerns the symmetry of the constructions involved. Whereas gapping constructions are almost invariably restricted to coordinate clauses with similar structures, the empty nominal heads are neither sentence-bound, in the sense that their antecedent has to occur in the same sentence (189), nor is the nominal phrase with the elided noun constrained to fulfill the same grammatical function as the one containing its antecedent (190).

(189) Hans hat sich *einen blauen Wagen mit vier Türen* gekauft,
Peter möchte gerne *einen* e *mit zwei Türen*.

(190) Hans hat *einen Wagen, der zwei Türen hat*, gekauft.
Der **e***, der vier Türen hat*, war ihm zu teuer.

In addition, in contrast to genuine gapping, an empty nominal head does not always require the presence of a linguistic antecedent, but can also be used deictically or in discourse situations where its reference can be resolved through the context. Utterances, such as (191), are perfectly acceptable in the appropriate context of a flower shop, for example.

(191) Ich hätte gerne die langstieligen gelben.

To summarize, it seems that none of the standard coordination ellipsis constructions can be directly related to our problem. The fact that there are some similarities should not be surprising, since in both classes of ellipsis some anaphoric relations are involved which concern syntactic and semantic aspects in varying degrees. While gapping appears to be most closely related to the nominal ellipsis, it is not at all obvious that an extension of the notion of gapping would contribute to solve the issue. In short, even though we cannot fully exclude the possibility of a more elegant solution eventually coming from this area, it appears equally safe and more adequate to go by an analysis which is based on the assumption of empty nominal heads.

Empty Nominal Heads and Further Constraints

If one decides to postulate an empty node functioning as an empty nominal anaphor, the main syntactic restriction which needs to be captured, next to those arising from anaphoric resolution, is to constrain the occurence of an empty head noun without any preceding modifiers or specifiers. The formal description of this restriction turns out to be a problem if one assumes that determiners as well as attributive adjectives are completely optional in constituents of a type *dp*.

We cannot simply require that a constituent of type *dp* must not be phonologically empty, since this constraint quite obviously fails to capture the crucial point.

A nominal phrase which consists of an empty head noun and some kind of complement to the noun or some post-modifier, such as a prepositional phrase or a relative clause (192), would satisfy the constraint of a phonologically non-empty *dp*. However, these constructions do not satisfy the constraint that the empty head noun has to be preceded by some phonologically non-empty material in a *dp*.

(192) *Hans mag *Wein, der süß ist*,
Peter bevorzugt **e**, *den man auch trinken kann.*

Similarly, we cannot capture the restrictions by assigning a specific type of category to the empty noun. For example, we could postulate that empty head nouns are of the type *np* and therefore obligatorily have to combine with a determiner or numeral. This would cover only the cases where we have some specifier (optionally followed by an adjective), but would not account for those cases where we have only an adjective combining with a mass noun or bare plural (193). In these latter cases the constraint on functional completeness of the maximal projection *dp* would be identified with the marking on the empty head and consequently lead to a contradiction.

(193) [Guter Wein] schmeckt, [schlechter **e**] verursacht Kopfschmerzen.

On the other hand, if we assigned the underspecified type *noun* to the empty head noun, so that it were identical to the category of mass nouns and bare plurals, the empty element could take all structural positions of these categories and thus lead to overgeneration again.

Note, by the way, that a solution along these lines would not even work, if we gave up the assumption that determinerless constructions are analyzed by means of underspecification and postulated an empty determiner instead. For example, one could define an empty determiner which takes as a complement a category of the type *np* and try to locate the constraints on the empty head noun in this lexical type. Such an approach would fail to achieve the desired effects if the emptiness of the head noun is encoded by means of a HEAD feature which can be accessed through a specification of an empty determiner as given in (194).

(194)
$$ {}_{\textit{empty-det}}\left[\begin{array}{ll} \texttt{LOC|CAT} & \left[\begin{array}{ll} \texttt{HEAD} & \left[\begin{array}{ll} \texttt{MAJOR} & \boxed{1} \\ \texttt{MINOR} & \left[\texttt{FCOMPL} \; +\right] \end{array}\right] \\ \texttt{SUBCAT} & \left\langle {}_{np}\left[\texttt{LOC|CAT|HEAD} \left[\begin{array}{ll} \texttt{MAJOR} & \boxed{1} \\ \texttt{MINOR} & \left[\begin{array}{ll} \texttt{FCOMPL} & - \\ \texttt{EMPTY} & - \end{array}\right] \end{array}\right]\right] \right\rangle \end{array}\right] \end{array}\right] $$

Such a specification would of course exclude all the ungrammatical constructions where an empty determiner is immediately followed by an empty head noun potentially followed by some post-modifiers, i.e., configurations of the sort (195).

(195) $[_{dp}$ empty-det $[_{np}$ empty-noun [mit Hut]]]

However, at the same time we would again exclude also those constructions where an adjective precedes the empty noun, since it is only the head information of the noun which becomes visible to the empty determiner through its subcategorization.

Similarly, we cannot obtain the required effects either, if we constrained the *np*-complement of the empty determiner to be phonologically non-empty. To begin, we have to keep in mind that the standard feature geometry of HPSG does not allow us to access phonological information through subcategorization, since the `SYNSEM` feature, which is accessible through the various selections features, does not include the `PHON` feature of a sign. But even if we rearranged the feature geometry to obtain access to the phonological string in the selection features, a constraint excluding phonologically empty complements would again be satisfied by the ungrammatical construction (195).

In short, alternative formulations of the restriction, requiring either a non-empty head or a non-empty phonology, will fail to provide the expected results, independent of the level of phrasal projection at which the constraints are applied. Thus, it does not suffice to encode the appropriate restrictions in the lexical category of the empty head noun nor can we exploit other already existing distinctions. Instead, additional devices will be necessary which allow to impose some structural constraints in addition to marking the empty element.

4.6.2 Non-empty Left Periphery of Nominal Phrases

In the following, we will discuss the different possibilities of capturing the relevant constraints sketched in the previous section by referring to the left periphery of nominal phrases. An analysis along these lines would basically postulate that an empty head noun must not occur as the left-most terminal node of a constituent of the type *dp*.

The main problem in this approach is that we have to refer to the linear order independent of the grammatical relations involved. In standard HPSG, there are only two devices through which direct reference to notions such as *left* or *right* daughter can be made. These devices are the Linear Precedence (LP) rules, which constrain the *possible* order of constituents, and the value of the `PHON` attribute, which encodes the resulting *actual* linear order of the `PHON` values of the daughter constituents. As a third device we will look at the introduction of some syntactic feature.

Linear Precedence Rules

Linear Precedence constraints, as they are traditionally understood, appear unsuitable to express the relevant restrictions for various reasons. To begin, traditional LP rules typically refer only to local configurations of depth one, whereas the required constraints would have to refer to the linear order of terminal nodes independent of the depth of embedding. Yet, even if one extended the notion of LP constraints to cover also non-local configurations, they would still not have the appropriate power; while LP rules are well-suited to express constraints over the linear order of existing constituents, they normally do not allow to enforce an existential quantification over a constituent. The LP rules informally sketched in (196) would only constrain the ordering of constituents relative to the empty noun if these constituents are existent, but none of the first two rules would enforce the presence of a determiner or adjective.

(196) det < empty-noun

adj < empty-noun

empty-noun < noun phrase

empty-noun < clause

More recently, however, a formalization of a language for LP constraints has been developed by [Manandhar 1995], which appears to allow for all of the required properties. This language comprises existential quantification and it permits to make reference to the first daughter in a sequence. Most importantly, LP constraints in this language are not constrained to configurations of depth one, but can have effect over arbitrary depths of embedding via an extension of the concept of *domain union*, originally developed by [Reape 1994]. The way to approach the problem in this language would be to form a domain union of the entire nominal phrase and then to impose a constraint at the top level of the *dp* that the first daughter must not be marked as empty.

Still, although it would be possible to express the necessary constraints in this language, the question arises whether domain union will be the appropriate and adequate device for this case. Domain union is normally employed to account for free word order relations by decoupling constituency relations, capturing grammatical dependencies, from linear precedence relations mirroring the surface order of terminals. In our case, the application of domain union operations would only have to serve the purpose of transporting information about the left-most terminal constituent to the maximal projection of a nominal phrase. At the same time, all other LP constraints governing the internal structure of the NP, which could be formulated equally well or even better at a more local level, would have to be maintained within the domain-unioned phrase. Again, this would not be a technical problem, but it appears questionable whether this device is not too powerful and whether simpler solutions (as we discuss them below) would not be more appropriate.

Constraints on PHON

The second strain of possible solutions we want to explore is based on reference to the PHON attribute, representing the phonological or graphemic structure of a linguistic sign by a sequence (string or list) of phonological or graphemic symbols. In this approach, the generalizations have to be captured by imposing filtering constraints over the PHON feature of structures to which LP constraints have already been applied. If we assume that the empty element is represented by an empty list $\langle\ \rangle$, the main problem in this approach is to identify this substring in the PHON sequence.

The most naïve and simplest approach of imposing a constraint on PHON feature of the *dp* itself to the effect that this value must not be prefixed by an empty list $\langle\ \rangle$ is of course doomed to fail.

(197) $$_{dp}\begin{bmatrix}\texttt{PHON}\ \langle\ \rangle \bullet \top\end{bmatrix} \quad \rightarrow \quad \bot$$

Since every list is equivalent to the concatenation of this list with an empty list, any list (empty or non-empty) will match the description of the antecedent.

The only possibility to formulate the constraint through the PHON feature thus appears to depend on having access to the composition of the resulting PHON sequence, as well as to the internal structure of the nominal phrase. The constraint over the left periphery could thus take the form of an implication (198) which not only refers to the mother node but also to the PHON value of the HEAD-DTR of the phrase.

(198) $$_{dp}\begin{bmatrix}\texttt{PHON}\ \boxed{1} \bullet \boxed{2} \\ \texttt{DTRS}\begin{bmatrix}\texttt{HEAD-DTR}\begin{bmatrix}\texttt{PHON} & \boxed{1}\langle\ \rangle \\ \texttt{SYNSEM} & _{noun}[\ ]\end{bmatrix}\end{bmatrix}\end{bmatrix} \quad \rightarrow \quad \bot$$

This constraint would filter out any constituent of the type *dp* as ill-formed, if the HEAD-DTR of this constituent is of the type *noun* and has an empty string as the value of its PHON attribute, and if this value is co-indexed with the prefix of the PHON value of the mother.

However, as it stands, this filter may well be as ineffective as the simple one given above, and if it is refined it could be the source of undecidability. Again, the problem is whether the antecedent in (198) sufficiently constrains the set of structures to which the filter is to be applied. In particular, it is questionable whether it will not also match a structure in which the empty string is not a prefix but an infix of the PHON string (199).

(199)
$$ {}_{dp}\begin{bmatrix} \texttt{PHON} & \boxed{1} \bullet \boxed{2} \bullet \boxed{3} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{HEAD-DTR} & \begin{bmatrix} \texttt{PHON} & \boxed{2}\langle\ \rangle \\ \texttt{SYNSEM} & {}_{noun}[\ ] \end{bmatrix} \end{bmatrix} \end{bmatrix} $$

To exclude this case, it would be necessary both to maintain all variable bindings occuring in the history of the derivation and to add an additional constraint, which excludes the occurence of the variable bound to the empty string in the HEAD-DTR, in the suffix of the mother's PHON attribute. This means that one may have to add a condition (200) involving negative list membership to the filter in (198).

(200) COND: $\boxed{1} \notin \boxed{2}$

Even if this constraint were still decidable, it will make the correct predictions only for those cases where the empty nominal head is immediately dominated by a root node of the type *dp*.

The constraint will fail to apply, however, when the head noun is empty but is not immediately dominated by a node of the type *dp*, as it would be the case with multiple post-modification (201).

(201) $*[_{DP}\ [_{N}\ [_{N}\ \mathbf{e}\ \ [_{PP}$ nach Paris $]\ ]\ ]\ [_{S}$ der kommt $]\ ]\ ]$

(202)
$$ {}_{dp}\begin{bmatrix} \texttt{PHON} & \boxed{1} \bullet \boxed{2} \\ \texttt{DTRS} & \begin{bmatrix} \texttt{HEAD-DTR} & \begin{bmatrix} \texttt{PHON} & \boxed{1}\langle nach, Paris\rangle \\ \texttt{DTRS} & \begin{bmatrix} \texttt{HEAD-DTR} & \begin{bmatrix} \texttt{PHON} & \langle\ \rangle \\ \texttt{SYNSEM} & {}_{noun}[\ ] \end{bmatrix} \\ \texttt{ADJ-DTR} & \begin{bmatrix} \texttt{PHON}\ \langle nach, Paris\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \texttt{ADJ-DTR} & \begin{bmatrix} \texttt{PHON}\ \boxed{2}\ \langle der, kommt\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} $$

To guarantee that the empty left periphery can be detected at any level of embedding, the filtering constraint would have to exclude that neither the top-level HEAD-DTR nor any HEAD-DTRs of a HEAD-DTR are empty heads. One possibility to express this generalization would be to employ functional uncertainty over an arbitrarily deep path of DTRS|HEAD-DTR and to require that none of the HEAD-DTR in this path must have an empty phonology.

$$(203)\quad \left[\begin{array}{ll} \texttt{PHON} & \boxed{1} \bullet \boxed{2} \\ \texttt{DTRS} & \left[\texttt{HEAD-DTR|(DTRS|HEAD-DTR)*} \begin{bmatrix} \texttt{PHON} & \boxed{1}\langle\ \rangle \\ \texttt{SYNSEM} & \textit{noun} \end{bmatrix}\right] \end{array}\right]_{dp} \rightarrow \perp$$

$$\text{COND: } \boxed{1} \notin \boxed{2}$$

Whether this expression is decidable is even more questionable as it involves universal quantification in combination with sort constraints.[31]

Of course, in some sense the entire problem would almost vanish into thin air, if we assumed that the empty nodes are not represented by an empty list or an empty string in the PHON value, but that there is some level of representation of the surface structure, in which *every* terminal element is reflected as an element in a sequence. Such a postulation is practically tantamount to the introduction of a new level of linguistic abstraction, since we would still need some representation (phonological or graphemic) which corresponds to the actual surface sound or string and which does not contain the symbols for the empty terminals anymore. Still, under these conditions, the very first and simplest approach becomes a possible option again, since we then can use a simple negative constraint referring to the first element of that sequence.

Left Periphery and Edge Features

There are some simpler solutions to the problem, however, if we abandon the idea of formulating the constraint directly through the PHON feature. In an alternative approach, whose spirit was first suggested to us by John Nerbonne, the relevant restrictions are captured on the basis of a syntactic feature which is percolated along the left-most daughters or edge of a phrase to its mother.[32]

Such a percolation mechanism would be already very effective if it was limited to those headed phrases in which the head is constrained to the type *noun*. For example, the definition in (204) specifies that if a HEAD-DTR of the type *noun* occurs as the left-most daughter somewhere within a headed phrase, then the value of an attribute LPE (for "left periphery empty") is identical to its value on the mother node.

[31] Undecidability was proven for the general class of constraints involving existential quantification inside universal quantification. Cf. [Baader et al. 1993].

[32] Cf. also [Nerbonne 1994a], who points out a similarity between this approach and the edge features suggested by [LaPointe 1990] and [Miller 1992].

(204)

$$
{}_{\textit{headed-phrase}}\left[\begin{array}{ll}
\texttt{PHON} & \boxed{1} \bullet \top \\
\texttt{SYNSEM} & \left[\texttt{LOC|CAT|LPE}\ \boxed{2}\right] \\
\texttt{DTRS} & \left[\begin{array}{ll}\texttt{HEAD-DTR} & \left[\begin{array}{ll}\texttt{PHON} & \boxed{1} \\ \texttt{SYNSEM} & {}_{\textit{noun}}\left[\texttt{LOC|CAT|LPE}\ \boxed{3}\right]\end{array}\right]\end{array}\right]
\end{array}\right]
$$

To obtain the appropriate effect we simply have to assign to the types of the empty noun and to the maximal projection *dp* two conflicting values, [LPE $+$] and [LPE $-$] respectively. These features will clash, if a *dp* is headed by an empty noun and if this head occurs as the left-most daughter of the phrase.

(205)

$$
{}_{\textit{dp}}\left[\begin{array}{ll}
\texttt{PHON} & \boxed{1} \bullet \boxed{2} \bullet \boxed{3} \\
\texttt{SYNSEM} & {}_{\textit{noun}}\left[\texttt{LOC|CAT|LPE}\ \boxed{4}\right] \\
\texttt{DTRS} & \left[\begin{array}{ll}
\texttt{HEAD-DTR} & \left[\begin{array}{ll}
\texttt{PHON} & \boxed{1} \bullet \boxed{2} \\
\texttt{SYNSEM} & {}_{\textit{noun}}\left[\texttt{LOC|CAT|LPE}\ \boxed{4}\right] \\
\texttt{DTRS} & \left[\begin{array}{ll}
\texttt{HEAD-DTR} & \left[\begin{array}{ll}\texttt{PHON} & \boxed{1}\langle\ \rangle \\ \texttt{SYNSEM} & {}_{\textit{noun}}\left[\texttt{LOC|CAT|LPE}\ \boxed{4} +\right]\end{array}\right] \\
\texttt{ADJ-DTR} & \left[\texttt{PHON}\ \boxed{2}\langle \textit{nach}, \textit{Paris}\rangle\right]
\end{array}\right]
\end{array}\right] \\
\texttt{ADJ-DTR} & \left[\texttt{PHON}\ \boxed{3}\ \langle \textit{der}, \textit{kommt}\rangle\right]
\end{array}\right]
\end{array}\right]
$$

There are some further phenomena however, which suggest that this percolation mechanism should be of a more general kind and that it should affect all left-most daughters in a nominal phrase, be they HEAD-DTRs or ADJ-DTRs.

These phenomena concern the distribution of empty head nouns in the context of a certain class of specifiers, above all the indefinite determiners *ein/kein* and the possessive pronouns. These determiners have two different and competing inflectional forms for the same slots in the CASE/NUMBER/GENDER paradigm. For the three singular slots of masculine nominative, as well as neuter nominative and accusative, we find forms without an inflectional ending like *ein*, *kein*, *mein* etc. as well as strongly inflected correlates *einer/eines*, *keiner/keines*, *meiner/meines* etc.

In contrast to other forms filling other slots in the paradigm these specific forms behave irregularly in that they have a complementary distribution with respect to their structural context in a nominal phrase. The uninflected forms are limited to a context where they are followed by either a pre-modifying adjective or a lexically overt head noun (206). The inflected forms on the other hand must *not* be followed by any of these (overt) categories (207).[33]

(206) ein Mann mit Hut
ein grosser **e** mit Hut
*ein / mein / kein **e** mit Hut

(207) *einer Mann mit Hut
*einer grosse(r) **e** mit Hut
einer **e** mit Hut
eines **e**

Basically the same phenomenon occurs with the genitive singular masculine and neuter forms of the definite determiner *des* / *dessen*.

(208) Ich habe mich des Mannes / *des **e** / *dessen Mannes / dessen **e** mit Hut erinnert

Although without this specific complementary distribution, there is yet another case where a complement of the type *np* is constrained to have a non-empty left periphery, namely prepositions with an incorporated dative determiner form (*am*, *zum* etc.).

(209) am ersten Tag nach Weihnachten
am zweiten **e** nach Weihnachten
*am **e** nach Weihnachten

In all of these cases post-modification does not have any effect on the grammaticality of the constructions whatsoever. Thus, for the same reasons as given above, we cannot simply generalize that the uninflected forms, the determiner *des*, or prepositions with an incorporated determiner should take only phonologically non-empty complements. These cases can be handled, if the *np* complements of these specific determiner forms were constrained to have a non-empty left periphery.

[33]It should be clear that for the inflected forms it is the positions in the paradigm corresponding to the uninflected forms which are relevant, not the morphological forms themselves. The forms are not subject to any such restrictions if they encode other CASE/NUMBER/GENDER configurations, as for example, genitive singular feminine (*einer Frau*) or neuter (*eines Kindes*).

(210)

$$
{}_{\textit{det-ein}}\begin{bmatrix}
\texttt{LOC|CAT} & \begin{bmatrix}
\texttt{HEAD} & \begin{bmatrix}
\texttt{MAJOR} & \boxed{1}\begin{bmatrix}\texttt{POS} & {}_{\textit{noun}}[\;]\end{bmatrix} \\
\texttt{MINOR} & \begin{bmatrix}\texttt{FCOMPL} & + \\ \texttt{SPEC} & +\end{bmatrix}
\end{bmatrix} \\
\texttt{SUBCAT} & \left\langle {}_{\textit{np}}\begin{bmatrix}
\texttt{LOC|CAT} & \begin{bmatrix}
\texttt{HEAD} & \begin{bmatrix}\texttt{MAJOR} & \boxed{1} \\ \texttt{MINOR} & \begin{bmatrix}\texttt{FCOMPL} & -\end{bmatrix}\end{bmatrix} \\
\texttt{LPE} & -
\end{bmatrix}
\end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Even under the restricted version of the percolation of the edge feature, this constraint will have the required effects and lead to feature clashes only in those situations where the complement *np* does not contain an adjectival modifier but has an empty head noun at its left periphery.

The more difficult problem is posed by those forms which must not be followed by a noun or adjective. Here, we basically have two choices.

If we want to exploit the `LPE` feature for handling these constructions, we have to generalize the percolation of this feature to all left-most daughters in a nominal phrase. The correct results are then obtained by constraining the complement of the inflected specifiers to have an empty left periphery. The lexical entry for the respective determiners would be identical to the one given in (210) except for the value of the `LPE` attribute. Given that all overt lexical items are then marked as [`LPE` $-$], this means that these determiners obligatorily combine with an empty head noun without any pre-modifiers but with an arbitrary number of post-modifiers.

Alternatively, we could reconsider the option of applying some nominalization process to these specifiers (and for reasons of consistency presumably to all other specifiers as well). This nominalization rule would take as input a determiner and eliminate the subcategorization for an *np*, but preserve all other properties, resulting in a structure such as (211).

(211)

$$
{}_{\textit{det-noun}}\begin{bmatrix}
\texttt{LOC|CAT} & \begin{bmatrix}
\texttt{HEAD} & \begin{bmatrix}
\texttt{MAJOR} & \begin{bmatrix}\texttt{POS} & {}_{\textit{noun}}[\;]\end{bmatrix} \\
\texttt{MINOR} & \begin{bmatrix}\texttt{FCOMPL} & + \\ \texttt{SPEC} & +\end{bmatrix}
\end{bmatrix} \\
\texttt{SUBCAT} & \langle\;\rangle
\end{bmatrix}
\end{bmatrix}
$$

This specification, which is equivalent to the definition of a *dp* except for the `SPEC` feature, will block any combination with a noun or some other nominal projection through complementation or with an attributive adjective through (pre-modifying) adjunction (due to the [SPEC $+$]) specification. However, a major problem for this approach is the treatment of post-modification, e.g., by means of genitive possessive phrases, prepositional phrases or relative clauses.

(212) [einer [meiner Großmutter]]
[einer [mit Hut]]
[einer [den ich gesehen habe]]

If all of these constructions are to be handled by some adjunction rule (rather than by complementation), almost the same problem as with the pre-modifying adjectives arises. The question is again how the (post-) modified category is to be specified. Again, it cannot be characterized as [FCOMPL $-$], since this would block determinerless constructions. If it is marked as [SPEC $-$], the above examples (212) would be excluded. However, if the modified constituent were constrained as [SPEC $+$] or not specified for this feature at all, overgeneration of expression such as (213) could arise, in which a post-modifier directly attaches to a determiner, before it combines with its complement.

(213) * [[die [meiner Großmutter]] Kekse]

The only option left open would be to leave both features FCOMPL and SPEC unspecified and to constrain the modified constituent as being a fully saturated constituent, which would block the derivation of (213). However, this also means that any number of spurious ambiguities have to be envisaged, since a post-modifier can now attach at both the *np* and *dp* level of projection. A more important effect would be that post-nominal possessives, which typically combine with nominal projections at the level of lexical items, cannot attach to unsaturated categories anymore, i.e., the combination of a post-nominal possessive and a prepositional or sentential complement would be excluded.

All of these problems do not arise if we maintain the systematic assumption of an empty head noun. Under this approach, the modified constituent in all post-modifying constructions can be marked by the constraint [SPEC $-$], which is satisfied by lexically overt or empty head nouns. The order between prepositional and possessive post-modifiers can be determined by restricting the possessives to adjoin to lexical categories only.

Now, of course it could be argued that these suggestions represent exactly what had been criticized in earlier sections, namely the introduction of empty terminals for the purpose of maintaining a more uniform structure. Yet, in contrast to empty determiners, which do not necessarily lead to a simplification of the analysis, it appears that empty head nouns are indeed empirically motivated and that they do support a more elegant and straightforward analysis of a range of different phenomena.

Pre-Specifiers

However, before we come to a final conclusion, we want to address an additional phenomenon which could turn out a potential problem to practically all of the solutions suggested above. These constructions appear to suggest that maximal nominal projections must have a non-empty left periphery, but that this condition by itself is not yet sufficient. Again, these constructions involve uninflected forms of categories such as *all* or *solch* as illustrated in (214) and (215).

(214) all die guten Kekse von seiner Mutter und
      * all **e** von seiner Großmutter

(215) Solch ein guter Wein aus Frankreich und
      * solch **e** aus Italien

At first sight, the constraints formulated above will fail to capture the ungrammaticality of these constructions, since the empty nominal head is preceded by some non-empty lexical material. Thus, the question arises whether the ungrammaticality of these constructions can be accounted for through some other generalization, which is independent of the issue of empty head nouns.

At least for the case of *all*, such a solution appears to be adequate. While the inflected forms of this lexeme is completely unrestricted (216), the uninflected forms appear to be bound to the presence of a determiner rather than the presence or absence of a head noun (217).

(216) alle diese guten Weine
      alle guten Weine
      alle **e**
      alle diese **e**

(217) all diese guten Weine
      all dieser gute Wein
      * all guten Weine
      * all Wein

This distribution can be accounted for if it is postulated that the uninflected *all* functions as an adjunct, which selects the lexical category of a definite determiner through its `MOD` feature. If such a determiner is missing, the pre-specifier cannot adjoin and the ungrammatical constructions in (214) and (217) are blocked. For the inflected forms, two different categorizations, as pre-specifying adjuncts and as specifier heads, could be assumed.

The situation is not quite as simple in the case of *solch*. Again, we find an inflected and an uninflected variant of this lexeme (218)/(219). However, this time the grammaticality of constructions involving the uninflected form appears to be independent of the presence of a determiner or a pre-modifying adjective, as the third example in (219) shows.

(218) ein solcher guter Wein
solcher Wein
ein solcher **e**
solcher **e**

(219) solch guter Wein
solch ein guter Wein
solch Wein
solch guter **e**
solch einer **e**
*solch **e**

Given that *solch* can also combine with determinerless constructions, the correct generalization could be that this form selects a maximal projection *dp*, which must have the semantics of an indefinite nominal phrase. The characterization of the selected constituent as *dp* will automatically require the left periphery of this phrase to be non-empty. Whether the selection is then based on complementation, where *solch* operates as a functional head requiring a functionally complete complement, or on adjunction to a full *dp* appears to be of secondary relevance.

4.6.3 Summary

In this section, we introduced some data involving missing head nouns. We argued against the assumption of some nominalization process or some ellipsis based on coordination reduction, and suggested the assumption of an anaphoric empty head noun instead. In the second part, we discussed several different approaches to capture the phenomenon that an empty head noun must not appear at the left periphery of a maximal nominal projection. Solutions which attempt to capture the restriction through the `PHON` attribute either require the introduction of some complex functional constraints (with unclear decidability status) or they presuppose the modification of the structure and nature of the `PHON` value.

Alternative solutions based on some syntactic relation between the mother and the left-most daughters appear to be considerably more straightforward, so we followed an approach in which a feature is percolated along the left periphery of nominal constituents. We discussed how such an assumption can be exploited to account for some complementarily distributed specifier forms, and argued that the consistent assumption of an empty head noun for practically all constructions with a missing noun will simplify the analysis of pre-nominal and post-nominal adjunction. We finally presented some apparent counterexamples of uninflected pre-specifiers which could be taken as evidence that a non-empty left periphery is only a necessary but not a sufficient condition for grammaticality, and showed how these examples can be accounted for by other generalizations, viz., adjunction to determiners and modification of maximal nominal projections.

4.7 Declension

In this section, we develop an approach to the distribution of declension classes in determiner–adjective–noun combinations. Our objective here is not only to provide an alternative description of the data but also to give some additional support for the assumption that determiners are best analyzed as functional heads of nominal phrases.

4.7.1 Declension Classes and Inflection

We base our analysis on three fundamental assumptions about the morphological paradigms of determiners, adjectives and nouns, and the morpho-syntactic specification of individual full-forms in the lexicon.

(i) There are only two declension classes in German, *strong* and *weak*, while the frequently postulated third *mixed* class can be dispensed with.

(ii) Determiners, adjectives *and* nouns are taken to be defined for a declension feature.

(iii) Uninflected determiners (and numerals) such as *ein* (in contrast to *einer*) are unspecified in the lexicon with respect to the declension class.

Since not all of these assumptions are fully uncontroversial, we motivate them in more detail in the following.

Elimination of the Mixed Paradigm

Assumption (i) crucially differs from more traditional approaches which organize the declension classes of adjectives (and nouns) according to the occupation of the specifier position preceding an adjectival form. In these approaches, adjectival forms occurring after determiners like *der, dieser* are assigned to the weak paradigm. Forms appearing in determinerless nominal constructions or following a prenominal genitive phrase are classified as strong. A third paradigm is postulated to account for the distribution of forms after a specific class of determiners including *ein, kein* and after the possessive pronouns. This paradigm, illustrated in (220), is called mixed because it is claimed to be composed of forms of the strong and weak paradigm.

We follow these traditional distinctions to the extent that we accept that there are competing inflectional endings in some slots of the case–number–gender paradigm, which are appropriately classified as strong and weak. With respect to the mixed paradigm, we follow those approaches which draw the line between strong and weak forms such that only the singular forms in the nominative masculine, and nominative and accusative neuter are marked as strong, whereas all other forms are assigned to the weak paradigm. In the table (220) the adjectival forms which we assign to the strong paradigm are italicized.[34]

(220) Mixed Inflectional Paradigm:

	mas		neu		fem	
Nom	*ein*	*alter*	*ein*	*altes*	eine	alte
Acc	einen	alten	*ein*	*altes*	eine	alte
Dat	einem	alten	einem	alten	einer	alten
Gen	eines	alten	eines	alten	einer	alten

The middle line in this table marks an alternative division between strong and weak, which is based on the assumption that the nominative and accusative forms belong to the strong paradigm, while the dative and genitive forms are classified as weak. The two classifications are descriptively equivalent since the remaining forms in the nominative (masculine and feminine) and accusative (feminine) case, marked in bold face below, coincide in the strong (221) and weak paradigm (222).

(221) Strong Inflectional Paradigm:

	mas	neu	fem	plural
Nom	alter	altes	**alte**	alte
Acc	**alten**	altes	**alte**	alte
Dat	altem	altem	alter	alten
Gen	alten	alten	alter	alter

(222) Weak Inflectional Paradigm:

	mas	neu	fem	plural
Nom	alte	alte	**alte**	alten
Acc	**alten**	alte	**alte**	alten
Dat	alten	alten	alten	alten
Gen	alten	alten	alten	alten

All by itself, this classification of forms and the elimination of the mixed paradigm is certainly not such a radical move. However, we intend to explicate this distinction ultimately on the basis of the general inflectional properties of the determiners

[34] The distinction is drawn in the same way by [Pollard and Sag 1994], citing also D. Wunderlich (p.c.). However, the explanation that they give differs considerably from the approach that we will outline below.

in whose contexts these forms occur, rather than stipulating that after an arbitrary class of determiners the adjectives have to be inflected weakly and after an equally arbitrarily defined class of determiners we find an even more arbitrary distribution of strong and weak forms.

Declension of Nouns and Determiners

The second assumption, viz., that not only adjectives but also nouns and determiners are assigned a declension feature, may need further motivating for nouns and determiners.

As for the first, there are clearly some nouns, such as *Verwandter, Beamter,* which show the same inflectional declension classes as adjectives (with some few exceptions). These lexemes have strong forms, which can be observed after uninflected determiners and after pre-nominal genitives (223), and they have weak forms occuring after inflected determiners (224).

(223) ein guter Beamter / Verwandter / Abgeordneter

(224) der gute Beamte / Verwandte / Abgeordnete

Most of these nouns have defective paradigms in that they do not have any neuter forms, which may be due to interferences with inherent, natural gender, and they typically do not have strong forms in the singular dative (see also below).

As to specifiers, it can be observed that all determiners (with the exception of some forms in the paradigm of the definite determiner *der*) and possessive pronouns can be morphologically decomposed into a stem and a paradigm of inflectional endings, which are almost identical with those in the strong adjectival inflection paradigm, e.g. *dies+er* (225). The obvious inflectional differences between the adjectival and determiner paradigms are limited to the genitive forms of the masculine and neuter singular (*-en* vs. *-es*). All other forms have either identical or closely related morpho-phonological realizations of the inflectional ending.

(225)

Mas	Nom	alter	der	dieser	(einer)	-er
	Acc	alten	den	diesen	einen	-en
	Dat	altem	dem	diesem	einem	-em
	Gen	alten	des	dieses	eines	-en/-es
Neu	Nom	altes	das	dieses	(eines)	-es
	Acc	altes	das	dieses	(eines)	-es
	Dat	altem	dem	diesem	einem	-em
	Gen	alten	des	dieses	eines	-en/-es
Fem	Nom	alte	die	diese	eine	-e
	Acc	alte	die	diese	eine	-e
	Dat	alter	der	dieser	einer	-er
	Gen	alter	der	dieser	einer	-er

Uninflected and Inflected Determiners

Next to the strong determiners listed in (225) above, there is also a small number of determiners (the indefinite determiners and the possessive pronouns), for which we find forms with strong and weak adjectival endings in different contexts, e.g., *dem einen Mann* vs. *einem Mann*. In particular, the existence of these forms with an explicit strong or weak ending leads to our third assumption, viz., that the forms *ein, mein, kein* are from a morphological point of view appropriately analyzed as stems without any inflectional ending and that these forms accordingly should not be assigned a value for the feature `DECL`, encoding the declension class.

We base this assumption on the postulate that the declension paradigms should be primarily established on the basis of morphological inflections rather than on the basis of structural contexts. In some sense, this assumption is but a natural consequence of the above assumption concerning the composition of mixed declension classes. This claim also represents the main difference between our approach and others which assume that declension paradigms have to be established on the basis of grammatical relations or structural contexts or conversely that the forms which appear in a particular structural context have to belong to the same declension class. However, although we assume that forms occuring in a specific structural context may cut across different declension classes, we also assume that the variations can be predicted on the basis of morphological contexts.

If we look at the table (225) above, the crucial differences are manifested in the paradigm of forms of the determiner *ein-* as well as related forms of the possessive pronouns or the determiner *kein-*. As it stands, this paradigm could not be established on the basis of combinations of determiner and noun, since then the slots taken by strongly inflected forms in parenthesis (*einer* and *eines*) would have to be filled by the uninflected form *ein* instead.

The forms in parentheses have inflectional endings which are all identical to the strong inflection of adjectival forms, but the only structural context in which these forms can occur are combinations with empty head nouns. Since all other forms can also occur in this specific context, one could of course assume that this structural environment provides the basis for establishing a strong paradigm for *ein-* and related specifiers.

The crucial question is, however, whether one should then assume in addition a second paradigm for *ein-* as a determiner of overt nouns which would be identical to the strong inflection in all but the three marked slots. Since there are no parallel inflectional paradigms, the postulation of this paradigm would have to be based exclusively on the structural context in which these forms occur. A paradigm established on this basis cannot coincide with an (existing) inflectional declension class, unless one either assumed a mixed declension class in which at least some forms are classified as weak, or even that all forms in this context should be marked as weak.

In both cases, at least the question arises what the assignment of the form *ein*, *mein*, etc., to a weak paradigm should be based on. This classification becomes a problem above all if one takes into account that there exists a full paradigm of forms for this class of specifiers in which the inflectional endings are identical to the weak adjectival paradigm (226)/(227).

(226) der *eine* alte Mann / der *eine* alte **e** / der *eine* **e**

(227) das alte Auto von Peter und das *meine* **e**

It has to be admitted that these forms are subject to some syntactic restrictions. The forms of *ein-* (226) have to occur after some definite specifier, the weak forms of the possessives can be used only in elliptical constructions with an empty nominal head (227). The negated specifier *kein* does not have a weak paradigm at all. However, it remains a fact that for the stems of these categories we find a strong and a weak declension paradigm with inflectional endings which are in both cases fully identical to the corresponding adjectival paradigms.

As a consequence we are faced with a situation where we have three different forms filling the same slot in the `CASE`–`NUMBER`–`GENDER` paradigms. For the nominative singular masculine slot for example we find the forms *meiner* – *meine* – *mein*. As long as we assume only two declension classes, there are always two competing forms for the same slot in the paradigm. Classifying *ein* as strongly inflected in complementary distribution to *einer* may help to solve this specific problem, but this would also cause havoc, if it comes to the description of the distribution of adjectival declension.

Thus, it would only remain to treat *meine* and *mein* as two competing weak forms. Yet, while the former could be classified as weak on the basis of two criteria, viz., an inflectional ending which is parallelled in the adjectival paradigm and occurence in a structural context which appears to be restricted to weak forms, neither of these criteria is fulfilled for the latter form. The assignment of *ein* and related forms to the weak paradigm cannot be based on a structural context generalization, because the other forms in that context are clearly strong forms. Nor can this assignment be based on a morphological analogy with other inflectional paradigms, since the null-ending only occurs with these forms.

One approach to account for this phenomenon could be to postulate two different categories which both have a weak inflectional ending, but which are realized by two different allomorphs (a null ending and -*e*). However, these allomorphs would be unusual in that they are not only bound to two different categories, but that they also trigger their morpho-syntactic context to be different. Whereas the inflected weak form triggers weak inflection on a following adjective (228), the allegedly weak uninflected form triggers strong inflection (229).

(228) der eine$_{we}$ alte$_{we}$ Mann

(229) ein$_{we}$ alter$_{st}$ Mann

One could argue, of course, that this different behaviour is caused by the fact that the determiner in the former case is an adjective and that adjectives always agree with each other, whereas the latter is a genuine determiner, which normally does not agree with following adjectives. However, if the inflected weak forms are classified as adjectives, they have to be marked as completely lacking a strong paradigm altogether, since constructions such as (230) are ungrammatical, but should be expected as determinerless constructions with *einer* functioning as an adjective.

(230) *einer$_{st}$ alter$_{st}$ Mann

There is yet another argument against the classification of *ein* as a weak form. By and large, the only motivation for this form to be classified as weak would have to come from the declension class of the following adjective. Such an analysis would have to be based on the assumption that the declension always changes between the determiner and the adjective, i.e., that a weak determiner requires strong declension, while a strong determiner requires weak inflection of the adjective. If this generalization were correct, it is quite unexpected that the combination of a weak determiner and a strong adjective does not occur with any other forms but the uninflected ones. Thus, one might expect that such a combination should also be possible with an inflected weak form of a determiner, as for example (231) for a dative feminine singular noun phrase.

(231) *einen$_{we}$ alter$_{st}$ Frau

Again, one would have to stipulate for no apparent reason that the paradigm of the indefinite determiner and possessives lacks the corresponding weak forms in all slots for which there is an existing strong form. In the comparison of our approach with the one taken by [Pollard and Sag 1994] in section 4.9, we will show that their approach, which is more or less based on such grounds, even stipulates such irregularities in addition for some forms of the otherwise perfectly regular inflectional paradigm of adjectives.

To avoid some of these complications and stipulations we pursue a different line of analysis which together with the assumptions made above will lead to a considerable simplification and straightforward description of the declension system.

Specifically, we postulate that the declension classes are more directly related to inflectional endings and that there is no allomorphy involved in realizing the same slot in an inflectional paradigm. Thus, we do not stipulate any inflectional *null*-morpheme for the uninflected forms and assign them neither to the strong nor to the weak paradigm, but assume that they are unspecified with respect to the declension class.

The forms *einer* – *eine* are accordingly specified as [DECL *st*] and [DECL *we*] respectively. The uninflected *ein* and related forms (including also uninflected numerals such as *zwei*) are not specified for a specific value for DECL at all and are assigned [DECL $\top$].

The distribution of the declension classes which we will describe in the following section will then be governed by general principles, which are partly structural in nature and partly lexical in the sense that they will have to make an occasional reference to certain classes of specifiers. However, they will not refer to specific forms of particular categories other than by the distinction between a straightforward strong and weak declension as well as inflected and uninflected forms.

4.7.2 Distribution of Declension Classes

Generalizations

Some of the data that we attempt to capture in this section have already been addressed in the introductory section and are illustrated once more in (232)–(237). In the examples we mark the declension class assigned to the forms by a subscript. We begin our discussion with some basic observations and generalizations.

In nominal constructions with a determiner, all the elements following the specifier are weakly inflected, if it is an inflected (definite or indefinite) determiner.

(232) die$_{st}$ lieben$_{we}$ Verwandten$_{we}$

(233) einem$_{st}$ lieben$_{we}$ Verwandten$_{we}$

If such a determiner is missing (234) or if we have an uninflected determiner or possessive pronoun (235), which has an unspecified declension value, the adjectives and nouns in such a construction are strongly inflected.

(234) liebe$_{st}$ Verwandte$_{st}$

(235) ein$_{\top}$ lieber$_{st}$ Verwandter$_{st}$

While determiners as the heads of a category *dp* and their complements of type *np* may have different marking, we always find agreement of the declension class within categories of type *np*, i.e., all adjectives and the nominal head must be assigned the same declension class.

(236) der$_{st}$ liebe$_{we}$ alte$_{we}$ Verwandte$_{we}$

(237) ein$_{\top}$ lieber$_{st}$ alter$_{st}$ Verwandter$_{st}$

These data can be described on the basis of the following assumptions. Declension classes are marked by means of the feature `DECL` with the values *st*, *we* and $\top$ for the strong, weak and uninflected forms respectively. Since determiners and their nominal complements can differ in `DECL`, this feature cannot be a `MAJOR` feature, which would be shared. Exploiting the government relation, which builds on the split into `MAJOR` and `MINOR` features, we treat the `DECL` feature as a `MINOR` property instead. This will account for the fact that functional heads and their complements may have different `DECL` values, just as they may have different `MINOR` specifications in general. The description of the above data can then be summarized in the following four generalizations:

(a) Categories of type *dp* are defined to be marked [`DECL` *st*].

(b) An inflected (strong) determiner requires the head of its complement to be marked [`DECL` *we*].

(c) An uninflected determiner (possessive pronoun, numeral etc.) shares the value of its `DECL` feature with its complement.

(d) Attributive adjectives agree in their declension class with the modified phrase selected through their `MOD` feature.

These generalizations not only account for the above basic distributions, but also capture the intuitions (a) that every nominal phrase has to bear some explicit, i.e., strong marking of the agreement features; (b) that in every phrase these features are encoded only once on the topmost head; (c) that determiners whose inflection does not satisfy such a requirement have to pass the constraint on to some subsequent forms; and (d) that all adjectives and nouns are on a par with respect to the explicit agreement features.

Formalization

The first of the above generalizations we can implement by making a final extension to the type definition of *dp*, requiring every constituent of that type to be marked [DECL *st*]. The effect of this specification is that the first head in a nominal phrase will be marked strong. As already mentioned, other morpho-syntactic agreement features such as NUMBER, GENDER and CASE are treated as head features under a MAJOR attribute AGR. Since the MAJOR features are shared between the head noun and the determiner anyway, the agreement of these properties will be established automatically through the definition of functional heads.

(238)
$$ {}_{dp}\left[\mathtt{LOC|CAT|HEAD}\left[\begin{array}{ll}\mathtt{MAJOR} & \left[\begin{array}{ll}\mathtt{POS} & \left[\begin{array}{ll}\mathtt{N} & +\\ \mathtt{V} & -\end{array}\right]\\ \mathtt{AGR} & {}_{agr}[\;]\end{array}\right]\\ \mathtt{MINOR} & \left[\begin{array}{ll}\mathtt{FCOMPL} & +\\ \mathtt{DECL} & st\end{array}\right]\end{array}\right]\right] $$

The second generalization describes the combinations with those determiners which show a (strong) inflection themselves, i.e., determiners such as *der, dieser, einem, meinem* (in contrast to the uninflected forms *ein, kein, mein* etc.) which lexically govern their complements to be strong.[35] Their structure is described by the type definition in (239).

(239)
$$ \textit{infl-det} \equiv \textit{det} \wedge $$
$$ \left[\mathtt{LOC|CAT}\left[\begin{array}{ll}\mathtt{HEAD} & \left[\mathtt{MINOR}\left[\mathtt{DECL}\ st\right]\right]\\ \mathtt{SUBCAT} & \left\langle {}_{np}\left[\mathtt{LOC|CAT|HEAD|MINOR}\left[\mathtt{DECL}\ we\right]\right]\right\rangle\end{array}\right]\right] $$

According to (c) all those functional heads in a noun phrase share their DECL value with their complements, which do not have an inflectional ending and are therefore unmarked for declension. We assume the uninflected forms of the indefinite determiners, the possessive pronouns, but also the uninflected numerals to be of that type. Together with the assumption made under (a), this has the effect that a nominal category of type *dp* will have strong declension imposed on its head noun and adjectives if it is headed by an uninflected determiner.

---

[35] Cf. [Zwicky 1986][968ff], who also analyzes declension as being best captured as a government relation. Even though there are quite a number of differences between the two approaches, our generalizations (a) and (b) could be argued to reflect his abstract distinction between *vertical* and *horizontal* government respectively.

$$(240)\quad \textit{uninfl-det} \equiv \textit{det} \wedge \left[\texttt{LOC|CAT}\left[\begin{array}{ll}\texttt{HEAD} & \left[\texttt{MINOR}\left[\texttt{DECL}\ \boxed{1}\right]\right]\\ \texttt{SUBCAT} & \left\langle {}_{np}\left[\texttt{LOC|CAT|HEAD|MINOR}\left[\texttt{DECL}\ \boxed{1}\right]\right]\right\rangle\end{array}\right]\right]$$

To account for the adjective–noun agreement, we have basically two choices. In the augmented type definition for attributive adjectives (241), the agreement relation between the adjective and the phrase to which it attaches is specified in a maximally non-redundant way directly as part of the selectional restrictions imposed by the `MOD` feature. In this approach it is assumed that the inflectional ending of the individual adjectives will be associated with values for `DECL` and `AGR` which are directly encoded in the `HEAD` feature of the adjective's `MOD` value.

$$(241)\quad {}_{adj}\left[\texttt{LOC|CAT|HEAD}\left[\begin{array}{ll}\texttt{MAJOR} & \left[\texttt{POS}\begin{bmatrix}\texttt{N} & +\\ \texttt{V} & +\end{bmatrix}\right]\\ \texttt{MOD} & \left\langle\left[\texttt{LOC|CAT|HEAD}\left[\begin{array}{ll}\texttt{MAJOR} & \left[\begin{array}{ll}\texttt{POS} & \begin{bmatrix}\texttt{N} & +\\ \texttt{V} & -\end{bmatrix}\\ \texttt{AGR} & {}_{agr}\left[\ \right]\end{array}\right]\\ \texttt{MINOR} & \begin{bmatrix}\texttt{DECL} & st/we\\ \texttt{SPEC} & -\end{bmatrix}\end{array}\right]\right]\right\rangle\end{array}\right]\right]$$

An equally satisfactory alternative is to define the adjective for the appropriate `HEAD` features and to express the agreement by means of coreferences.

(242)

$$\begin{bmatrix} \mathtt{LOC|CAT|HEAD} & \begin{bmatrix} \mathtt{MAJOR} & \begin{bmatrix} \mathtt{POS} & \begin{bmatrix} \mathtt{N} & + \\ \mathtt{V} & + \end{bmatrix} \\ \mathtt{AGR} & \boxed{1}\ {}_{agr}[\ ] \end{bmatrix} \\ \mathtt{MINOR} & \begin{bmatrix} \mathtt{DECL} & \boxed{2}\ st/we \end{bmatrix} \\ \mathtt{MOD} & \left\langle \begin{bmatrix} \mathtt{LOC|CAT|HEAD} & \begin{bmatrix} \mathtt{MAJOR} & \begin{bmatrix} \mathtt{POS} & \begin{bmatrix} \mathtt{N} & + \\ \mathtt{V} & - \end{bmatrix} \\ \mathtt{AGR} & \boxed{1} \end{bmatrix} \\ \mathtt{MINOR} & \begin{bmatrix} \mathtt{DECL} & \boxed{2} \\ \mathtt{SPEC} & - \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}_{adj}$$

This latter analysis appears to be preferable, however, if one wants to maintain a cleaner separation between morphology and syntax. For example, if it is postulated that morphology produces as output a set of morpho-syntactic features which are to be linked only to the `HEAD` features of a lexical item, this could be satisfied by the latter, but not by the former approach. The situation we encounter here is basically the same as the one with subject–verb agreement, which could also described either as a government or as an agreement relation (cf. section 3.3 above). In both cases, we would prefer an agreement approach, since it supports the definition of the interface between syntax and (inflectional) morphology in a locally more constrained way, even if it may require an additional coreference to link the `HEAD` features with the selection features.

Examples

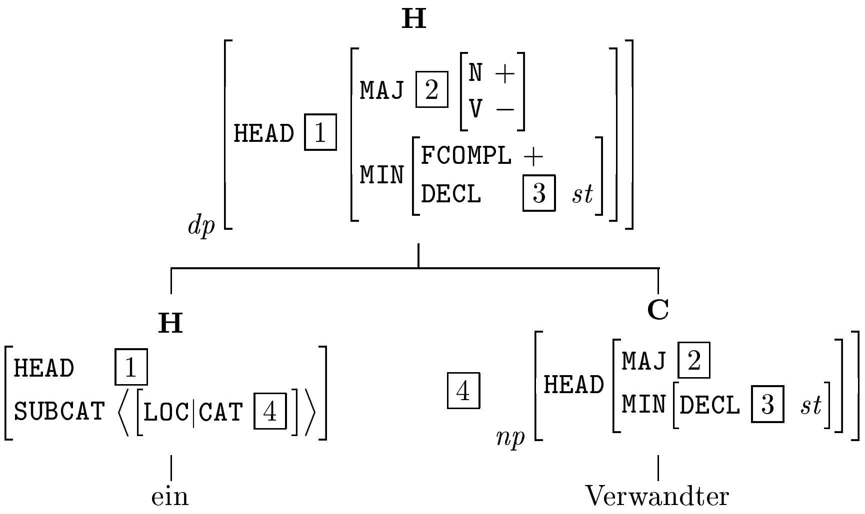
It is easy to see how the above definitions account for the distribution of the `DECL` values in the different determined and determinerless constructions. The analysis crucially draws on the fact that the determiner as the functional head of the nominal phrase can function as a mediator between the declension restriction on the maximal projection *dp* and the `DECL` value of the determiner's complement by controlling the flow of information in the `MINOR` feature. If such a determiner is absent the restrictions have to be fulfilled directly by the nominal head of the phrase. Any adjective adjoining to this head has to agree with respect to the declension and agreement features.

In the simplest cases, where we have a bare plural noun which shows inflectional variation of the declension class, e.g. *Verwandte*, only the strongly inflected morphological forms are compatible with the type definition for a *dp*, whereas others have to combine with an inflected determiner first.

(243)
$$
{}_{dp}\begin{bmatrix}
\texttt{PHON} & \langle \textit{Verwandte} \rangle \\
\texttt{SYNSEM} & \begin{bmatrix} \texttt{LOC|CAT} & \begin{bmatrix} \texttt{HEAD} & \begin{bmatrix} \texttt{MAJOR} & \begin{bmatrix} \texttt{POS} & \begin{bmatrix} \texttt{N} & + \\ \texttt{V} & - \end{bmatrix} \\ \texttt{AGR} & \begin{bmatrix} \texttt{NUM} & \textit{pl} \end{bmatrix} \end{bmatrix} \\ \texttt{MINOR} & \begin{bmatrix} \texttt{FCOMPL} & + \\ \texttt{DECL} & \textit{st} \end{bmatrix} \end{bmatrix} \\ \texttt{SUBCAT} & \langle \ \rangle \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

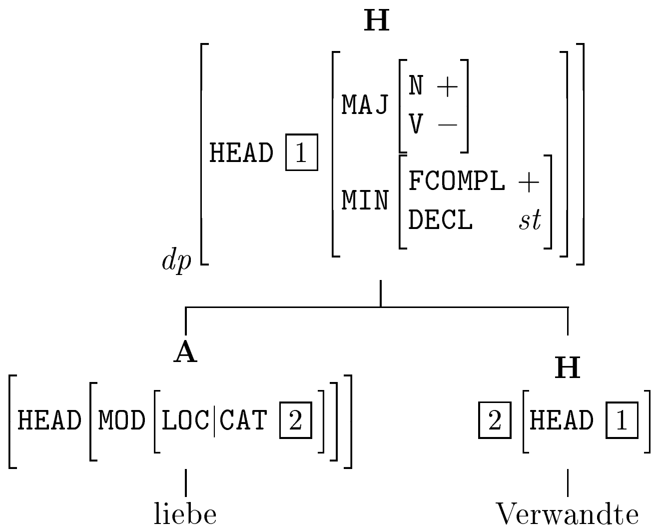
If a noun is combined with a determiner there are two possibilities. If the determiner form belongs to the (strongly) inflected class, it will assign [DECL *we*] to its *np* complement resulting in a form such as *der Verwandte*. If the determiner form is not marked for a DECL value it will receive the specification [DECL *st*] through the definition of the type *dp*. According to the definition of uninflected determiners it identifies its DECL value with the one of its *np* complement, with structures such as (244) for *ein Verwandter* as a result.

(244)



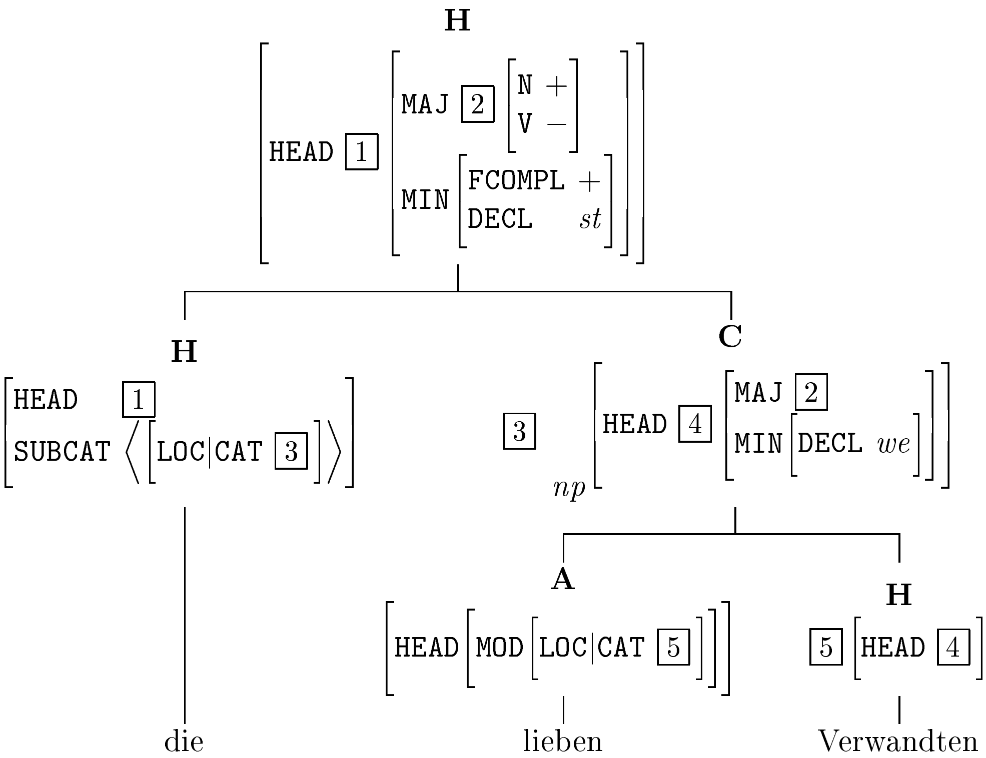
If there is no determiner at all (245), the assignment of DECL will be analogous to the bare plural construction in (243) above, consisting of a noun only. The root node of the maximal *dp* constituent is here expanded through a *head-modifier-schema*. The constraint requiring strong declension of the head percolates from the *dp* mother node directly to the noun, which is selected by the adjective through its MOD feature. Through the interdependence between the adjective's MOD and HEAD features adjunction to the noun is constrained to the appropriate adjectival inflection forms.

(245)



In the case of determiner–adjective–noun constructions (246), the determiner will first assign a DECL value to the HEAD-DTR of its complement. Through the HFP this value is percolated along the projection path headed by the noun. Any adjective attaching to this projection path will be again constrained via its selection feature MOD, such that it will only combine with a phrase with the appropriate DECL value.

(246)



Open Problems

Before we go on to show in the next section how our approach carries over to an analysis of numerals in pre-nominal specifier positions, we want to point out some apparently idiosyncratic exceptions of adjective–noun agreement which are outside the scope of our assumptions.

These exceptional cases arise from certain morphological variations in the dative singular forms, where sometimes only the first element in a series of adjectives carries the strong marking while all the other forms are weakly inflected (247). The same construction in the nominative (or any other) case (248) consistently shows a strong inflection on all forms.

(247) Roms oberstem$_{st}$ neuen$_{we}$ Abgeordneten$_{we}$

(248) Roms oberster$_{st}$ neuer$_{st}$ Abgeordneter$_{st}$

Even if we find variants of these constructions where all adjectives bear strong dative marking, these examples are not easily accounted for by any of the known approaches. The crucial issue is again that there is basically no way in which this first element in a row of adjectives can be singled out, as long as adjunction is treated by the standard approaches.

One possibility could be to shift the locus of the treatment of declension from a syntactic level to a morphological level. It could be assumed, for example, that `DECL` is a property of a separate level of morphology which represents the morphological properties of the terminals by means of a list of feature structures and which can be accessed through selection mechanisms. In this case, the lexical entries of the respective strong adjective forms could impose a constraint on the constituents selected through their `MOD` feature requiring the first morphological element in the sequence to be marked as weak. The major flaw with this approach is that it does not take into account that an adjectival head in a prenominal adjunct does not always have to be the first element. In complex adjectival phrases such as (249), this approach would fail, because the first terminal element of the modified constituent need not be the head of the modifier, but rather can be any other category.

(249) Roms oberstem$_{st}$ [neu gewählten$_{we}$ Abgeordneten$_{we}$]

The information required to handle this sentence appears to involve at the same time head feature percolation (inside the modifying phrase *neu gewählten*), more constrained locality features, as well as access to the linear order, including notions like 'first element in a sequence of a certain type'.

A more promising approach would be to follow the alternative strain of analyzing prenominal adjectives as heads or functors, as we briefly sketched it above in our discussion of categorial-grammar-style analyses for adjuncts. Under this assumption, DECL would have to become a local feature so that it is not shared between the modifier and its argument. The strong dative form of an adjective would then simply require its argument to be marked weak. All other forms in the adjectival paradigm would of course continue to show agreement or token identity of the DECL value with respect to their complements.

However, even this approach is not fully satisfactory, since it does not account for the fact that this type of construction is just one possibility. While it is difficult enough to handle the case were only the first adjective is marked as strong (250), it is even more difficult to account for the second option at the same time, viz., that in all other cases the strong marking must be present on all modifying adjectives (251) vs. (252)

(250) Gutem$_{st}$, frischen$_{we}$, französischen$_{we}$ Wein

(251) Gutem$_{st}$, frischem$_{st}$, französischem$_{st}$ Wein

(252) *Gutem$_{st}$, frischem$_{st}$, französischen$_{we}$ Wein

Since the adjectival head cannot see beyond the head of the constituent it governs, at least two features would be necessary to handle these cases on a level of lexical selection or government. Thus, we have to admit that we are well aware of this problem but also that we do not have a straightforward solution to it at the moment.

4.7.3 Summary

In this section we first outlined some basic assumptions on the morphological distinctions we make in terms of declension classes. We argued that the assumption of a third, mixed declension class is redundant, not only because it is composed of the two other classes, but mainly because there is no common denominator which unifies this class in an explanatory way.

We suggested instead to establish and treat declension classes not on the basis of the structural context involving certain lexical items (such as the indefinite determiners or possessives), but rather on the basis of the morphological forms of these lexical items. In this sense, we rejected an analysis in which a mixed declension is determined and established by the respective phrases being specified by the determiner *ein-* etc. Instead, we account for these declension phenomena on the basis of the distinction between inflected and uninflected governing specifiers. In particular, we showed that the relevant class of determiners exhibits three different competing declension forms in certain slots, viz., strong, weak and unmarked.

Given these basic assumptions, we could describe the phenomena on the basis of some few, very simple generalizations, all of which are both maximally non-redundant and encode the apparent idiosyncrasies where they actually occur in the paradigm. None of these axioms has any effects which are not restricted to a local context, which are not lexically based, and which have to take into account any information outside the paradigm of the lexical item itself. As we will argue in the comparison with other approaches (section 4.9), not all of these properties can be claimed for the approach by [Pollard and Sag 1994].

4.8 Numerals

Since the descriptive generalizations as well as some of the simple examples presented quite obviously could also be handled by other approaches, we will show in this section, how our generalizations determining the distribution of declension classes will also carry over to more intricate cases involving numerals such as (253) and (254):

(253) zwei$_\top$ liebe$_{st}$ Verwandte$_{st}$

(254) meine$_{st}$ liebsten$_{we}$ zwei$_\top$ Verwandten$_{we}$

4.8.1 Some Basic Data

Under the category of numerals we generally subsume uninflected forms of cardinals such as *zwei* or *einhundertein* but also inflected forms such as *beide* or *einhunderteiner*. These categories and lexical items, which will be in the focus of our discussion, appear to share properties with both modifiers and specifiers.

Numerals have in common with adjectives as pre-nominal modifiers that they can appear in the complement position of determiners (255), that they can be more or less freely permuted with adjectives (256), and that they have to agree with respect to the DECL value of preceding and following adjectives or nouns (256) vs. (257). This agreement is evident in the case of inflected numerals (*beide*), while it can be stipulated in the case of uninflected numerals from the fact that preceding and following forms have to have the same weak declension.

(255) die *zwei* neuen Beamten
die *beiden* neuen Beamten

(256) die neuen *zwei* Beamten
die neuen *beiden* Beamten

(257) *die neuen *zwei* Beamte
*die neuen *beide* Beamte
*die neuen *beiden* Beamte

On the other hand, numerals also behave like specifiers insofar as they can take the positions of determiners and that the inflected forms can govern the DECL value on the head of their complements (258). Uninflected numerals, just like uninflected determiners, do not affect the strong declension of the following constituents (259), unless they are themselves governed by a determiner.

(258) *beide* neuen Beamten
**beide* neue Beamte

(259) *zwei* neue Beamte

The notable exception to this generalization are the genitive forms of the numerals ending in *zwei* or *drei* for which we find a strongly inflected form. Adjectival forms after these numerals have to show strong inflection.

(260) *zweier* neuer Beamter
**zweier* neuen Beamten

The lexeme *beide* has an additional determiner-like property insofar as it can also be used as a pre-determiner, modifying or specifying another determiner.

(261) *beide* diese neuen Beamten

(262) **zwei* diese neuen Beamten

A particularly intricate complication arises through the fact that numerals cannot follow an adjective with a strong inflection (263).[36] This presents a problem insofar as numerals clearly can be marked for strong declension, if they behave similar to determiners (261)/(262), while they can also appear in an adjectival position, if they have a weak inflection (256).

(263) *neue *zwei* gute Beamte
*neue *beide* gute Beamte
*neuer *zweier* guter Beamter

As a last restriction, which could potentially also be captured on a semantic level, it should be mentioned that normally only one cardinal can occur in a noun phrase.[37]

(264) *die drei neuen zwei jungen Beamten

4.8.2 Numerals as Adjectives or Functional Heads

The most crucial issue to be solved in the analysis of numerals concerns the evidence for treating numerals as functional heads or rather as modifiers along the lines of adjectival adjuncts. For example, the fact that the forms of *beide* go together with weak declension on the following adjectives, could be taken as some strong evidence for treating this lexical item as a functional head. However, the question is whether this evidence suffices to treat simple cardinals also as heads rather than adjuncts.

[36] These examples were brought to our attention by Reinhard Karger.

[37] John Nerbonne pointed out to us that it would be problematic to induce semantic restrictions for these expressions, if one takes into account that there are constructions which do not show any semantic anomalies or contradictions but are nevertheless ungrammatical.

(i) *die mindestens drei neuen höchstens zwei jungen Beamten

As one of the main criteria for classifying a category as a (nominal) functional head, we assumed so far that its combination with an inherently functionally incomplete category qualified the constituent as a maximal projection. However, as long as we do not assume that the forms of *ein-* are also classified as cardinals, this criterion is neutralized, since numerals typically combine with a plural form, which could function as the head of a maximal projection even if there were no specifier.

Thus, the reliable test case or evidence of a construction, in which a singular count noun qualifies as a maximal projection if the determiner is replaced by a cardinal, is practically excluded by the nature of numerals.[38] As an exception one could only induce those cases where a complex cardinal combines with a *singular* count noun, as in *Tausendundeine Nacht*. However, this evidence could be countered by the argument that all forms ending in a suffix *-undein_* should be reduced to a coordinative construction. This assumption is corroborated by the comparison with the variants of cardinals without the conjunction *und*.

(265) Frau S. erzählte tausend und eine Nacht / *Nächte lang Geschichten.

(266) Frau S. erzählte tausendeine *Nacht / Nächte lang Geschichten.

What remains are those cases where numerals appear to have a different distribution than ordinary adjectives, as for example the constructions showing that numerals cannot be embedded under (strongly inflected) adjectives if there is no other determiner, as well as the cases illustrating the restrictions on the multiple occurrence of numerals.

However, even if we cannot fully and convincingly reject alternative approaches, we will pursue and outline an equally adequate analysis below in which numerals are treated as functional categories. Our analysis will then be based on the following assumptions:

(a) Numerals are functional heads. As such they only share their `MAJOR` features with their complements.

(b) Like all functional heads numerals take a functionally incomplete category as their complement. They have in common with determiners that this category has to be of the type *np*.

[38]This is not to say that the cardinality denoted by cardinals stands in a direct correlation to the number occuring on the noun, as the counterexamples below or cases such as (i) show.

(i) Null Flaschen_{pl} stehen_{pl} auf dem Tisch.

Note by the way, that a semantic treatment of the plural agreement, as suggested by [Pollard and Sag 1994], could be somewhat difficult in this case.

(c) Numerals differ from determiners insofar as they themselves are not marked for functional completeness. Thus, they do not automatically project into a constituent of the type *dp*.

(d) In order to block multiple occurences, cardinals are assigned a `MINOR` specification [`CARD` +] with the restriction that their complement is marked as [`CARD` −].

The first assumption (a) can be seen as a more or less arbitrary stipulation, the others are natural consequences of the analysis of numerals as heads. The second claim, that numerals take complements of the type *np*, reflects the observation that numerals cannot take a constituent as a complement which is headed by a determiner. (In an analysis of numerals as adjectival adjuncts this would fall out from the fact that they have to modify a constituent which is marked [`SPEC` −].)

Assumption (c) introduces a new dimension into the analysis of functional heads, in that it decouples the classification of a category as a functional head from the property of functional completeness. This assumption contrasts with the definition of the type of functional categories *func-cat* in section 4.4 above as being inherently functionally complete. On the other hand, such a move opens up the possibility that a phrase may contain more than one functional head, specifically that a numeral can become the head of a functionally complete or incomplete category of the type *dp* or *np* respectively.

Crucially, these three assumptions will also allow us to express the restriction in (d) as a syntactic constraint, viz., that the *np* complement of a cardinal must not contain another embedded cardinal as its (functional) head. Switching the value of the `CARD` feature in the `MINOR` head features will thus have the same effect as changing the `FCOMPL` feature in determiner-noun constructions, enforcing the uniqueness of the respective category of functional heads.

Note, that the option to constrain numerals to unique occurences on a syntactic level is practically excluded in an analysis of numerals as adjuncts. As for any other selectors, any constraints which a modifier imposes on the modified constituent on the basis of some `LOCAL` feature have no effect outside this configuration, since the information does not disseminate beyond the strictly local domain, unless some specific percolation mechanism were stipulated. However, if the restriction were on the head features of the modified constituent, the effect would quite obviously be neutralized just as well, since modifiers of the same type cannot impose conflicting constraints on the same head. In short, since modifiers are selectors which can change or affect the properties of their arguments with respect to the resulting phrase only on the semantic level, the singularity of numerals as adjuncts could only be captured through some semantic feature.

To formalize the analysis of numerals as functional heads, we can either redefine the subsumption hierarchy in section 4.4, such that the type *func-cat* does not inherit from *fcompl-cat* anymore, or we can introduce a generalized supertype for numerals and categories of the type *func-cat* which abstracts away from this property. In the latter case the only information to be added in the type description of cardinals will be the specification of the `MINOR` feature `CARD`.

(267)
$$ {}_{card}\left[\texttt{LOC|CAT}\ \left[\begin{array}{ll} \texttt{HEAD} & \left[\begin{array}{ll}\texttt{MAJOR} & \boxed{1}\\ \texttt{MINOR} & \left[\texttt{CARD}\ +\right]\end{array}\right] \\ \texttt{SUBCAT} & \left\langle {}_{np}\left[\texttt{LOC|CAT}\ \left[\texttt{HEAD}\ \left[\begin{array}{ll}\texttt{MAJOR} & \boxed{1}\\ \texttt{MINOR} & \left[\begin{array}{ll}\texttt{FCOMPL} & -\\ \texttt{CARD} & -\end{array}\right]\end{array}\right]\right]\right]\right\rangle \end{array}\right]\right] $$

Through a second set of assumptions we capture the distribution of declension classes in constructions with numerals. These generalizations do not only exploit the assumption that numerals can be analyzed as functional heads, but also follow to a large degree the more general principles determining the distribution of `DECL` outlined in the previous section.

So, for example, the generalization under (e) is basically a repetition and extension of the scope of our assumptions above concerning uninflected determiners like *ein* etc.

(e) (*preliminary*) The `DECL` value of uninflected numerals is token identical to the `DECL` value of the heads of their complements.

The effect of (e) is of course the same as if we had analyzed numerals as adjectival adjuncts, which simply agree and do not change any declension values.

Similarly, the second assumption (f), which is intended to cover the case of *beide*, is practically identical with our generalizations about inflected determiners, which we claimed always to govern weak inflection.

(f) (*preliminary*) Inflected numerals require their complements to be weakly inflected.

If the numeral itself is weakly inflected (*die* $neuen_{we}$ $beiden_{we}$ $Beamten_{we}$), the assumption under (f) is of course observationally indistinguishable from an agreement relation (described by token identity of values). Nevertheless, this formulation of the rule is simpler, since it does not have to distinguish two cases. Instead of two generalizations, viz., (i) a strongly inflected *beide* governs weak marking on the complement, and (ii) weak inflection agrees with the weak inflection of the complement, we have only one general rule.

Although these generalizations would nicely corroborate our approach to declension, the two assumptions in this form are not quite satisfactory insofar as they do not cover the exceptional case of the genitive forms *zweier* and *dreier*, while they do exclude the (obsolete) forms of *beide* agreeing with their complement's declension (268).

(268) beider / zweier / dreier alter Männer

The most direct way to account for these constructions would be to drop the qualification *uninflected* in (e) and simply assume that all numerals can agree with their complements. If one wants to include the obsolete *beide* constructions, the generalization in (f) would have to be made an optional condition restricted to *beide*.

(e) The `DECL` value of numerals is token identical to the `DECL` value of the heads of their complements.

(f) The forms of *beide* behave according to (e) or require their complements to be weakly inflected.

It should be noted that these formulations to some degree appear to argue against the broad generality of our assumptions on inflected and uninflected determiners above. If we wanted to maintain these more important generalizations, we could always accept the alternative analysis of numerals as adjuncts and simply exempt the *beide* constructions. Under this assumption, uninflected numerals of course would not introduce additional evidence for the analysis of uninflected determiners. However, the case of the inflected forms like *zweier* would not run against the analysis of inflected determiners, while the *beide* constructions could still be accommodated.

Even then, the generalizations, that we gave so far, account for most of the data presented above, but do not yet capture the restriction that only the weak forms of numerals, in particular *beide*, can be embedded under adjectives.

Here again, we cannot seek for a solution based on the dimension of functional completeness, if we want to exclude structures such as (269)/(270), where an adjective precedes the (strongly inflected) numeral in a determinerless construction.

(269) *junge$_{st}$ beide$_{st}$ neuen$_{we}$ Beamten$_{we}$

(270) *junge$_{st}$ zwei$_{st}$ neue$_{st}$ Beamte$_{st}$

To block these constructions, we have to resort to the MINOR feature SPEC again, which does play a restricting role in adjectival adjunction. What appears to be required is a covariation of the DECL values with the possible values of SPEC signalling the presence or absence of a specifier. If a numeral is marked as strong, it must also be marked as [SPEC +], i.e., it must not be modified by a preceding adjective. For the weak declension forms the negative SPEC value holds. In the extended type definition of numerals (271) this correlation is expressed by distributed disjunctions over the values of the features DECL and SPEC.

(271)
$$ {}_{\mathit{card}}\left[\texttt{LOC|CAT|HEAD|MINOR}\left[\begin{array}{l}\texttt{SPEC}\ \$1\{+,-\}\\ \\ \texttt{DECL}\ \$1\{st, we\}\end{array}\right]\right] $$

Through this specification the ungrammatical examples (269)/(270) are blocked. In these constructions, the specification [DECL *st*], introduced through the type *dp*, will percolate down to the numeral as the functional head, where it will trigger the first disjunct of the distributed disjunctions to be chosen, i.e., [SPEC +]. Since adjectives are constrained through their MOD values to attach to constituents marked [SPEC −], a preceding adjective will cause a feature clash. If on the other hand an adjoining adjective enforces a specification [SPEC −] on the numeral, the weak declension has to be chosen, leading to a contradiction at root of the *dp* if the maximal projection is headed by the determiner.

We can assume that the specification in (271) holds for all cardinals. This class is further divided into two subtypes (272) and (273) which represent the inflected and uninflected forms respectively.[39]

(272)
$$ {}_{\mathit{infl\text{-}card}}\left[\texttt{LOC|CAT}\left[\begin{array}{ll}\texttt{HEAD} & \left[\texttt{MINOR}\left[\begin{array}{l}\texttt{SPEC}\ \$1\{+,-\}\\ \texttt{DECL}\ \$1\{st, we\}\end{array}\right]\right]\\ \\ \texttt{SUBCAT} & \left\langle {}_{np}\left[\texttt{LOC|CAT|HEAD|MINOR}\left[\texttt{DECL}\ we\right]\right]\right\rangle\end{array}\right]\right] $$

[39]If we take the inflected forms of *zwei* etc. into account, (272) will only hold for *beide*, while (273) will hold for all other cardinals.

(273)

$$ {}_{\textit{uninfl-card}}\begin{bmatrix} \texttt{LOC|CAT} & \begin{bmatrix} \texttt{HEAD} & \begin{bmatrix} \texttt{MAJOR} & \boxed{1} \\ \texttt{MINOR} & \begin{bmatrix} \texttt{DECL} & \boxed{1}\ \$1\{st, we\} \\ \texttt{SPEC} & \$1\{+,-\} \end{bmatrix} \end{bmatrix} \\ \texttt{SUBCAT} & \left\langle {}_{np}\begin{bmatrix} \texttt{LOC|CAT|HEAD} & \begin{bmatrix} \texttt{MAJOR} & \boxed{1} \\ \texttt{MINOR} & \begin{bmatrix} \texttt{FCOMPL} & - \\ \texttt{CARD} & - \\ \texttt{DECL} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} $$

4.8.3 Summary

In this section, we explored how our approach to determiners as functional heads carries over to the treatment of numerals, which could be alternatively analyzed as adjectival modifiers. We showed that an analysis as functional heads can be accommodated to a large degree with our approach, if we give up the assumptions that all functional heads must be marked as functionally complete. This position would have to be abandoned anyway, if one wants to allow multiple functional heads within a construction. As we showed by the example of numerals, a co-occurence of several functional heads of the same type can be easily blocked through restrictions in the `MINOR` head feature.

The second part of the section was concerned with the declension phenomena. It turned out that a formulation of the appropriate restrictions would be both more complicated and less satisfying, if we strictly follow previous assumptions about inflected and uninflected specifiers. Restricting the generality of these assumptions leads to a more straightforward description of the relevant phenomena. It should be kept in mind, however, that the classification of numerals as functional heads nevertheless facilitates the analysis of constructions with diverging declension classes as well as of phenomena illustrating a covariation between declension class and specifier role.

4.9 Comparison and Summary

In the following, we will compare the approach we proposed in this chapter with the analysis suggested in [Pollard and Sag 1994]. It should be pointed out, though, that large parts of our analysis had been developed as reactions to draft versions of [Pollard and Sag 1994]. This is relevant insofar as some of the assumptions in the body of [Pollard and Sag 1994] eventually had been revised in the well-known final chapter 9, which in some sense can be seen as indicating a move from "HPSG II" to "HPSG III". Most of our criticism will still hold even against this revised versions. However, there are also some revisions, above all those concerning the encoding of selection in separate valency features, which will make the difference between our approach and some current HPSG assumptions appear less drastic. In our comparison, we will concentrate on the slightly older approach and indicate wherever improvements had been suggested by [Pollard and Sag 1994].

As we already pointed out above (e.g., section 3.3) the [Pollard and Sag 1994] analysis of specifiers, adjectival adjunction and declension in nominal phrases is based on the assumption of a mutual selection of noun and determiner. We will present this approach and show that it differs from our own in three crucial aspects: (i) the treatment of determinerless constructions, (ii) the analysis of declension phenomena and (iii) the issue of adjectival adjunction. The difference with respect to the first aspect could largely be reduced to a technical contrast revolving around the opposition between disjunction and underspecification. The differences here have diminished in the revised version of chapter 9. The treatment of declension in [Pollard and Sag 1994], on the other hand, necessitates a much more indirect analysis, which might also lead to conflicts in the empirical predictions. Finally, we will show that the problem of adjectival adjunction, as we described it in 4.5, will also lead to overgeneration in the analysis of [Pollard and Sag 1994]. We will suggest an appropriate modification to this theory, which—even if it does not reduce the complexity of the analysis—at least eliminates the source of overgeneration.

4.9.1 Treatment of Specifiers in Pollard and Sag 1994

In their analysis of nominal phrases and specifiers, [Pollard and Sag 1994] assume that nouns and determiners mutually select each other by means of two different selection mechanisms. In spite of the presence of two functors in one local configuration, this approach is claimed to represent a "conservative" solution in the NP-analysis tradition, since the noun continues to function as the head of a nominal phrase, while the determiner has full access to all information contained in the $\overline{\mathrm{N}}$-constituent which it specifies.

The head property of nouns in this analysis manifests itself unambiguously through the fact that the noun (and its modified projections) is assigned the function of a `HEAD-DTR` and thus provides the `HEAD` features in nominal phrases. The relation between the noun and the determiner is treated as a head-complement relation, where the noun determines the presence or absence of a specifier through its `SUBCAT` list.

(274) $$\begin{bmatrix}\texttt{SYNSEM} & {}_{noun}\begin{bmatrix}\texttt{LOC}|\texttt{CAT} \begin{bmatrix}\texttt{SUBCAT} \left\langle {}_{det}[\ ] \right\rangle\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

In the revised version, the selection of the determiner through the noun is represented by a specific valency feature `SPR` which serves to select specifiers only. While so far this approach is equivalent to a pure NP-analysis, it is also acknowledged by [Pollard and Sag 1994] that the determiner should be able to select the projection of the head noun, mainly to ensure a correct semantic treatment of quantifiers. Since the noun already selects the determiner through the `SUBCAT` (or `SPR`) feature and since a phrase cannot have two different heads, the selection of the noun by the determiner is handled by an additional, new selection mechanism. Determiners (275) are assigned a special `HEAD` feature `SPECIFIED`, short `SPEC`, which takes as its value the `SYNSEM` value of the head which subcategorizes for the determiner.

(275) $$\begin{bmatrix}\texttt{SYNSEM} & {}_{det}\begin{bmatrix}\texttt{LOC}|\texttt{CAT} \begin{bmatrix}\texttt{HEAD} & \begin{bmatrix}\texttt{SPEC} \ {}_{noun}[\ ]\end{bmatrix} \\ \texttt{SUBCAT} & \langle\ \rangle\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

The phrasal combination of these categories is shown in (276) for a nominal phrase consisting of a determiner and a noun, where the noun and the determiner have access to all each other's `SYNSEM` properties through the `SUBCAT` and `SPEC` features respectively.

(276)

$$\begin{array}{ccc}
 & \textbf{H} & \\
 & \begin{bmatrix}\texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \langle\ \rangle\end{bmatrix} & \\
\textbf{C} & & \textbf{H} \\
\boxed{2}\begin{bmatrix}\texttt{HEAD} & \begin{bmatrix}\texttt{SPEC}\begin{bmatrix}\texttt{LOC}|\texttt{CAT}\ \boxed{3}\end{bmatrix}\end{bmatrix} \\ \texttt{SUBCAT} & \langle\ \rangle\end{bmatrix} & & \boxed{3}\begin{bmatrix}\texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \left\langle\begin{bmatrix}\texttt{LOC}|\texttt{CAT}\ \boxed{2}\end{bmatrix}\right\rangle\end{bmatrix} \\
| & & | \\
\text{Det} & & \text{N}
\end{array}$$

One aspect of this structure, the subcategorization of the noun for the determiner, is covered by the standard *head–complement–schema* (Schema 1 in the original version) and the SUBCAT Principle. The coindexation of the head daughter with the value of the complement daughter's SPEC attribute has to be defined through a new mechanism, which is introduced as a SPEC Principle:[40]

(277) SPEC Principle
If a non-head daughter in a *headed–phrase–structure* bears a SPEC value, it is token identical to the SYNSEM value of the head daughter.

In the revised version the basic assumptions of a mutual selection are maintained, however a new schema, the *head-spr-schema*, is introduced which is practically identical to the above configuration in (276) except for a replacement of the SUBCAT feature by the more specific valency feature SPR. It does not become clear, however, whether the SPEC Principle is still maintained or whether its effect is hard coded into the new schema.

While the treatment of adjectival modification in German is based on the standard *head-adjunct-schema*, it is suggested to solve the specific problems of agreement and the distribution of declension classes by some rather intricate extensions. The central idea of the approach is that an adjectival form of a certain declension class not only selects a noun through its MOD feature, but also imposes constraints on the declension class of the determiner that the noun has to be subcategorized for. For example, in the structure of *der alte Wein*, the weak form of the adjective *alte* would select a noun that subcategorizes for a strong determiner, e.g., *der*:

(278) $$_{adj}\left[\texttt{LOC|CAT|HEAD}\left[\texttt{MOD}\ _{noun}\left[\texttt{LOC|CAT}\left[\texttt{SUBCAT}\ \left\langle {}_{det\text{-}strong}[\ ] \right\rangle\right]\right]\right]\right]$$

In chapter 9, this approach is replaced by an alternative solution which acknowledges that declension is a property that could easily be assigned to nouns as well as to adjectives. Under this assumption declension can then be treated as a head feature of the noun, such that the adjective could in principle either agree with the noun in its declension class or alternatively select nouns with specific declension classes. This assumption forms the basis for a revision of the selectional relations. Thus, determiners are not assumed to be selected by the adjective directly anymore, but rather select the noun through their SPEC features. On the basis of an

[40] It should be noted that these specifications go beyond earlier version of HPSG in [Pollard and Sag 1987] by explicitly introducing cyclic feature structures, which arise through the mutual subcategorization of determiner and noun in (276). More significantly, while the antecedent of the SPEC Principle may look fairly harmless, it appears to require substantial logical or meta-logical extensions. This is at variance with most other principles which appear to fit smoothly into the underlying feature logic.

opposition which runs very much along the line of our distinction between inflected and non-inflected determiners, it is assumed that inflected determiners select weak nominal phrases, while uninflected determiners, including also numerals such as *zwei*, will select nominal constituents with strong declension.

4.9.2 Comparison with Pollard and Sag 1994

In the following, we will compare the two versions of the [Pollard and Sag 1994] approach outlined above with our own. We will concentrate on the issues of determinerless constructions, declension and adjectival adjunction.

Determinerless constructions

In the discussion of a pure NP-analysis (section 4.2), we already addressed one of the main differences between the basic [Pollard and Sag 1994] approach and our own, viz., that the contrast between determined and determinerless constructions under standard assumptions has to be captured either by means of disjunctive subcategorization or through the postulation of empty determiners, rather than by means of underspecification. Thus, it appears that in order to derive constructions such as *die Kinder* and *Kinder*, mass and plural nouns have to be subcategorized disjunctively for a determiner or for an empty list of complements:

(279) $$\left[\texttt{SYNSEM}\ {}_{noun}\left[\texttt{LOC|CAT}\left[\texttt{SUBCAT}\ \left\langle {}_{det}[\ ]\right\rangle \lor \langle\ \rangle\right]\right]\right]$$

Alternatively, one could maintain a uniform subcategorization for all nouns and derive determinerless constructions by means of an empty determiner. Nouns which obligatorily require a determiner would subcategorize for a non-empty determiner, others would subcategorize for a determiner, but leave it unspecified whether it must be overtly realized.

As SUBCAT is only replaced by the feature SPR in the revised version, chapter 9 of [Pollard and Sag 1994], basically the same holds also for this approach. However, since the SPR feature is only accessed and referred to in one specific schema, the *head-spr-schema*, there is theoretically also the possibility to leave the lexical entries of the respective nouns unspecified. In this case the selection feature SPR would have to be introduced in the structural rule as part of the HEAD-DTR specification, such that this rule could either apply or not. Ignoring the fact that under this assumption not all selector feature are rooted in the lexicon, this would make our two approaches almost equivalent. However, as we will see below, the split of head features and the treatment of the specifier as head, may still lead to a simpler solution in other respects.

Declension

The second difference concerns the treatment of declension in the two approaches, specifically the direction of selection or government of declensional paradigms, and the locality of the domains of the involved selection mechanism.

In our approach, the direction of government leads from the maximal nominal projection, which is defined as being *strong*, via the determiner (or alternatively head noun), to the NP complement, from where it can be accessed through the adjective's `MOD` selection. In contrast, the path of selection in [Pollard and Sag 1994] starts at the adjective and leads via the noun's subcategorization to the determiner. In the simplest cases, this means that we claim that a strong determiner requires its complement noun and all adjoining adjectives to be weak, while [Pollard and Sag 1994] assume that a weak adjective (indirectly) requires a strong determiner by selecting a noun which subcategorizes for such a determiner. Similarly, where we assume that a determinerless construction is required to carry strong declension on the head noun and adjective through the definition of a maximal *dp* projection, [Pollard and Sag 1994] assume that a strong adjective requires the head noun to have a saturated `SUBCAT` list, i.e., that it must not select a determiner anymore.

This difference also manifests itself in an extended domain of locality of selection for the approach by [Pollard and Sag 1994]. While in our analysis the specification of the selectional constraints imposed by the individual selectors is restricted to the `HEAD` features (or other non-selection features) of the selected category, the analysis of adjectival modifiers in [Pollard and Sag 1994] does not obey such limitations. This is noteworthy, since in most other cases selection of a category in HPSG is restricted to the non-selectional local features or involves at most the degree of saturation of the selected category.[41]

In part, such locality restrictions were intentionally 'hard-wired' by the reorganization of the feature architecture in the transition from [Pollard and Sag 1987] to [Pollard and Sag 1994], which shifted the type of the values of selection features from *sign* to the type *synsem* (and in some cases even smaller structures). Apparently, the intention was to constrain the selection of a category to information which does not go beyond the strictly local domain, and above all to block any reference to phrase structure nodes outside such strictly local configuration. As we already pointed out (section 3.3), it is technically (i.e., in terms of feature geometry) not at all easy to enforce such a limitation under all circumstances, as long as selection involves features whose values contain some recursion mirroring more extended structural domains. In particular, it would be practically impossible to

[41]The notable exception to this restriction is the analysis of *easy* adjectives. Cf. [Flickinger and Nerbonne 1991] and [Pollard and Sag 1994][166ff].

enforce a comparable locality restriction for the case at hand without imposing overly strict limitations and giving up well-motivated flexibility. However, even if there are no such technical solutions to the problem, the question arises whether one should not declare such a restriction to be a methodological constraint, specifically while there are perfectly valid alternative solutions available.

A further closely related aspect, in which the two approaches crucially differ, is the treatment of—in our terminology—uninflected determiners or possessives (*ein, mein, kein*, etc.). For these forms we assumed that they are underspecified with respect to the declension feature in the lexicon, that they receive their value either from a governing head or through the specification of the root node, and that they pass the value on to their complements. In principle, this meant that an uninflected determiner could be marked as strong or as weak, depending on the context in which it appears. In contrast, [Pollard and Sag 1994] stipulate that these forms are explicitly marked as belonging to the weak inflectional paradigm and that the relevant strong forms of the adjectives would select these determiners via the noun's SUBCAT feature.

Since it is only specific forms in the strong adjectival paradigm which combine with these particular determiners, this analysis has as a consequence that the exceptional behaviour and inflectional marking of some determiners have to be accounted for in the otherwise fully regular adjectival paradigm, rather than being located in the paradigm of the determiners, where the idiosyncrasy actually occurs. While all other strongly inflected adjectives select a noun with an empty subcategorization list, the strong forms in the singular, nominative masculine, and nominative and accusative neuter, i.e., the two forms *alter* and *altes* in their relevant usages, have to be marked as selecting a noun with a disjunctive subcategorization for no determiner or a weak determiner.

(280)
$$_{adj}\left[\texttt{LOC|CAT|HEAD}\left[\texttt{MOD}\ _{noun}\left[\texttt{LOC|CAT}\left[\texttt{SUBCAT}\ \left\langle {}_{det\text{-}w}[\ ]\right\rangle \vee \langle\ \rangle\right]\right]\right]\right]$$

This specification not only involves a redundant encoding of irregularities found in the paradigm of indefinite determiners and possessive pronouns (since this irregularity is duplicated in the adjectival paradigm); it also exposes once more a certain lack of methodological restrictions on the locality of specifications. Under this approach the inflectional paradigms of adjectives and determiners are not defined uniformly and independently, but rather the definition of one paradigm (the adjectival one) has to take into account the possible structural contexts in which the forms may occur as well as the regularities and above all irregularities of other paradigms.

Of course, much of this criticism does not hold for the revised version of chapter

9 anymore, since there the direction of the selection of the declension class is the same as in our approach. The entire nominal constituent selected by the determiner is marked for a certain declension, and thus the differences in selecting different declensions can be encoded in the individual forms of the determiners. However, it appears that even this analysis suffers from some shortcomings and inadequacies which could be solved in our approach.

Finally, while our generalizations on determiners easily carry over to the treatment of inflected or uninflected numerals, it is not easy to see how constructions such as (281) – (284) could be handled without further complications within the approach of [Pollard and Sag 1994]:

(281) zwei$_{st}$ gute$_{st}$ Freunde

(282) meine$_{st}$ zwei$_{we}$ guten$_{we}$ Freunde

(283) beide$_{st}$ guten$_{we}$ Freunde

(284) meine$_{st}$ beiden$_{we}$ guten$_{we}$ Freunde

In our analysis we would treat both *zwei* and *beide* as functional heads. For the uninflected cardinals, we predict that they are uninflected and therefore unmarked for declension in the lexicon, but that they are assigned a feature value for declension through the structural context: they are marked *strong* if they occur as the topmost functional head of a nominal phrase (281), *weak* if they are governed by a strong determiner (282). Since these forms co-index their value for declension with their complement's declension, the following adjectives have to be strong and weak respectively. *Beide* on the other hand, is an inflected functional head and always requires its complement to be marked weak, while it may itself be marked as *strong* as the topmost head (283) or as *weak* if governed by another inflected functional head (284).

This rather simple approach, where uninflected numerals show 'agreement' whereas inflected numerals govern with respect to declension, requires different assumptions in the original version of [Pollard and Sag 1994]. As far as we can see the most consistent strain of analysis in that approach would be to assume that *zwei* does not function as a specifier at all, but rather as an adjunct, which is unmarked for declension in both constructions (281) and (282). In the first case the adjective would be strong, because there is no specifier; in the second case it is weak, because *meine* functions as a specifier. Treating cardinals as adjuncts means, of course, that one has to face the problem of multiple occurences again, which have to be excluded on the semantic level.

For *beide* on the other hand, quite different assumptions have to hold. For the strong form of *beide* in (283), a specifier function has to be assumed, so that it can be selected by the noun (and indirectly by the adjective). Whereas it would be a simple modifier showing agreement with the adjective in the second case (284).

Even if one followed the revised version of chapter 9, the picture is not entirely coherent and homogeneous. The assumption here is that inflected determiner and numerals always select a nominal constituent which is marked weak. This accounts for the case of *beide*, provided that the occurence of two specifiers can be accommodated with the approach. However, the approach still fails in the case of the uninflected cardinals, which are assumed to select strong nominal constituents, contrary to the evidence of (282).

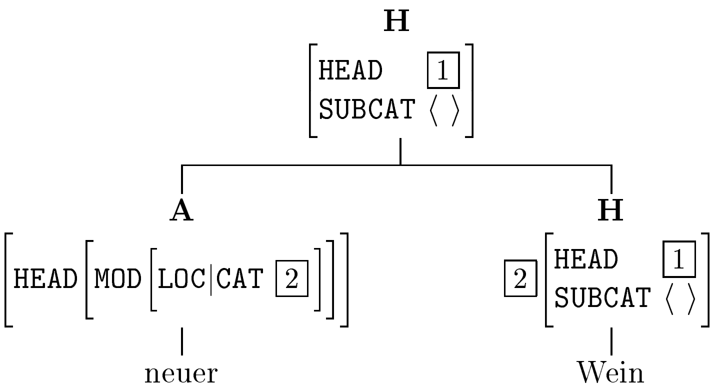
Adjectival adjunction

The third and presumably most crucial difference between the two approaches concerns the treatment of adjectival adjunction. While our approach covers the problem on a purely lexical basis, the analysis of [Pollard and Sag 1994] requires different types of solutions, ranging from a change of the principles and basic assumptions of the theory along some lines which we will suggest below, via the obligatory postulation of empty determiners, to the introduction of a new schema—all of which are purely structural solutions.

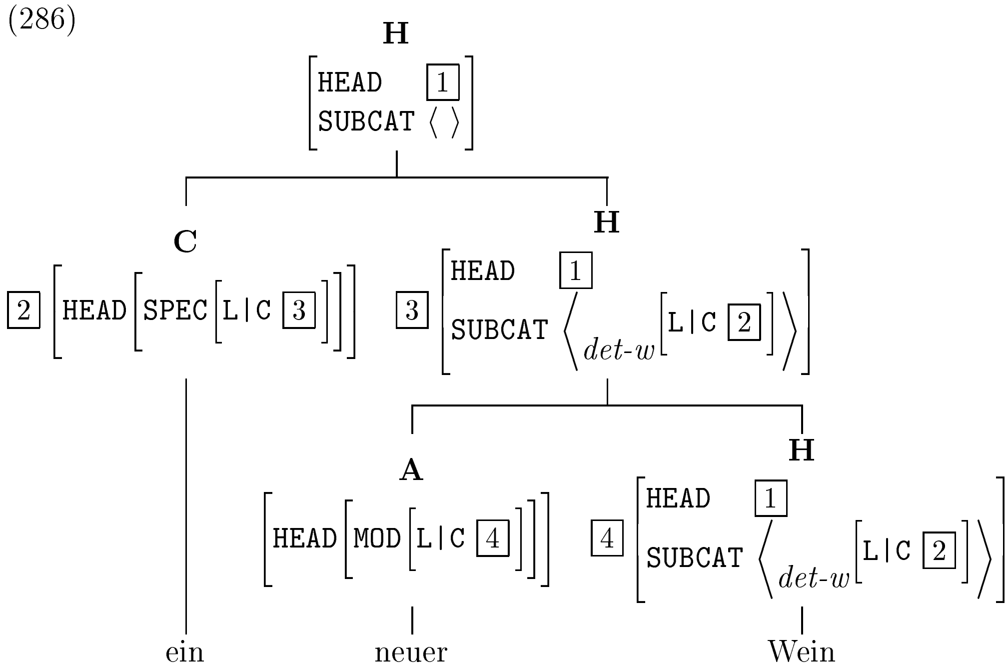
In the original version of [Pollard and Sag 1994], a specific and idiosyncratic subclass of the strongly inflected adjectives select a noun with an empty subcategorization list through their MOD feature. This assumption together with the fact that there is no difference between a SUBCAT list which was always empty and which has been saturated does not allow the blocking of derivations such as *alter ein Wein* without further change to the theory.

We illustrate the problem with three different examples involving the adjective *alter* with strong declension. The two wellformed structures (285) and (286) show the adjective selecting a noun which subcategorizes for no determiner and a weak determiner respectively. (Cf. also the entry for these adjectival forms in (280) above.)

(285)



(286)

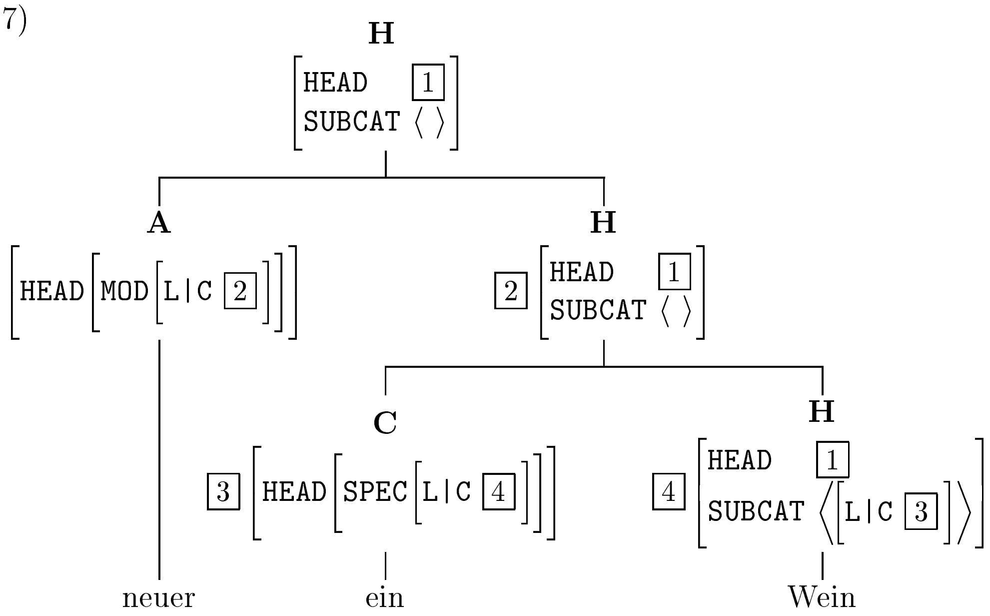


As is easy to imagine, the first use of the adjective, in which it selects a noun with an empty subcategorization, is again the source of over-generation. In the ungrammatical structure (287) the adjective is adjoined to a nominal projection in which the subcategorization of the noun for a determiner has already been satisfied, yielding as a result an empty `SUBCAT` list.

We can safely conclude that saturation of the `SUBCAT` list is as ill-suited for defining the intermediate levels of projection, which are relevant for adjectival adjunction, as functional completeness turned out to be in our approach (section 4.5). An empty `SUBCAT` list may have arisen through the head noun never having required a specifier, or it may have resulted from a previous combination of a noun with a subcategorized-for complement.

However, in contrast to our approach, the analysis of [Pollard and Sag 1994] does not provide a way to handle the problem in the lexicon by appropriately marking the entries of determiners, adjectives or nouns. The reason for this is simply that there is no direct relationship or flow of information between a determiner and its mother node which is not mediated by the head noun. To block adjunction to a phrase containing a determiner, a specifier has to be able to mark its presence on the mother of this phrase, which can only be achieved through the sister node with which is stands in various selectional relations. Yet, the only principled connections between a head and its mother node are through the `HEAD` and the `SUBCAT` feature. Since the latter can be neglected in this case, the determiner would have to signal its presence through the head features of the noun. This

(287)



information would then be percolated up to the mother node by virtue of the HFP. However, it would also be percolated *down* within a phrasal head. As a result adjectival adjunction would be blocked everywhere, including also the grammatical example (285).

On the other hand, a lexical solution is feasible in our approach simply because in our analysis there are two different *head domains*, one headed by the determiner and the other one headed by the noun. Even though these two domains crucially share certain information, viz., the HEAD features encoded under the MAJOR attribute, they still differ in their MINOR features. This allows us to project the required information from the lexical entry of the determiner to its mother node through the standard HFP.

Since in the analysis of [Pollard and Sag 1994] there is only one head domain—the one headed by the noun—the only properties (besides SUBCAT) that are by definition 'non-local' in the sense of being shared between different phrasal projections are the entire set of HEAD features. Thus, any selectional constraint that a specifier or a modifier imposes on a head is therefore either local and not projected to the mother, or it is shared between *all* projections of the head through the HFP.

Of course, one could again opt for a structural solution to the problem on the basis of empty determiners or a unary structural rule.[42] As an alternative we sketch in

[42][Cooper 1990], who attempts to integrate $\overline{\text{X}}$-theory into HPSG and who defines different levels of projections independent from subcategorization, claims to be able to block adjunction of adjectives to maximal projections. Granted an extension of his underlying feature logic, this is certainly correct. However, he also admits (p.c.) that he will need some additional device

the following a more conservative extension for the theory of [Pollard and Sag 1994] which relies on shifting some information from a rule schema to a principle. The schema that we build on is introduced by [Pollard and Sag 1994][44ff,51] to cover the related phenomenon of complementizer constructions (288).

(288) (Schema 4) a phrase with `DTRS` value of sort *head-marker-structure* whose marker daughter is a marker with a `MARKING` value token identical to that of the mother

The introduction of this additional schema is motivated by the assumption that determiners are selected through subcategorization by head nouns, while there is no such relation between complementizers and verbs. Determiners and complementizers have in common, however, that they both carry a `SPEC` feature, which triggers the `SPEC` Principle and the selection of the head through these categories.

Drawing on this commonality between determiners and complementizers, the solution builds on generalizing that part of Schema 4 which establishes a direct structural connection between the complementizer and its mother node through the `MARKING` feature. By eliminating the clause concerning the `MARKING` attribute from Schema 4 (289), and shifting it to a place which affects all specifiers, viz., the `SPEC` Principle (290), a connection between all specifiers and their mothers will be established.

(289) (Schema 4 revised) a phrase with `DTRS` value of sort *head-marker-structure*

(290) `SPEC` Principle (revised)
If a non-head daughter in a *headed–phrase–structure* bears a `SPEC` value, it is token identical to the `SYNSEM` value of the head daughter and the non-head daughter's `MARKING` value is token identical to that of the mother.

Given this formulation of the `SPEC` Principle, all specifiers, no matter whether they function as complement daughters or as marker daughters, are able to signal their presence on the mother node. Determiners are then specified in the lexicon in the same way as complementizers as having a feature [`MARKING` +]; while adjectives require the head to which they adjoin to be specified [`MARKING` −]. This latter constraint could be of an even more general nature, since adjunction to a phrase containing a complementizer appears to be also ungrammatical. If this assumption turned out to be correct, a more appropriate locus of the negative `MARKING` restriction could be the head daughter in the general *head-adjunct-schema*.

While this approach provides a quite viable and general solution to the problem of adjunction, it still has to be kept in mind that it crucially builds on the `SPEC`

along the lines just mentioned to be able to derive specifierless constructions.

Principle, which is not only a structural principle but also one of the more complex principles of the theory.[43]

The simplest solution (in some sense) thus appears to be the choice of introducing an entirely new schema, as it is the case in the revised version of chapter 9, where the *head-spr-schema* is proposed as a schema next to the *head-marker-schema*. Since such a move represents a structural solution *par excellence*, the marking of the presence of a determiner should be trivial.

Suffice it to say that our approach to the problem can easily be generalized to the treatment of complementizers, as we will show in the next chapter. This allows us to dispense with the SPEC Principle as well as with the schemata for head-complementizer and head-specifier structures altogether.

4.9.3 Summary

In this chapter we developed a new approach to functional categories, which differs from others insofar as it does not declare either a functional or substantive category as the exclusive head of a maximal projection. With respect to nominal phrases, we tried to show in section 4.2 that such an assumption will lead to problems and complications (i) for the analysis of phrases without an overt head, viz., determinerless constructions or constructions with an empty nominal head, (ii) for the treatment of certain phenomena which in some way or other involve head features, as for example declension classes, and (iii) for the definition of non-maximal projections, as it is required for an adequate treatment of adjectival adjunction.

We therefore introduced a notion of functional heads in section 4.3, which not only accounts for a series of empirical observations, but also provides a theoretical basis with enough flexibility to solve the above as well as a number of other problems. This notion crucially builds on the assumption, that phrases can be structured into (at least) two distinct head domains, one headed by a functional head, the other by a substantive head. The head domain of the substantive head is assumed to be governed by the functional head, but in the absence of a functional head the substantive head may function as the head of the maximal projection itself, if it satisfies certain requirements. The different domains share certain properties, which are declared as MAJOR head features, while they may differ in others, viz., those which are introduced under a MINOR head feature. We introduced the basic definitions for substantive and functional categories and introduced a notion of functional completeness, which serves to define maximal projections in combination with the property of argument saturation.

[43] Cf. footnote 40 above.

In section 4.4, we showed how the abstract type definitions for substantive and functional properties can be naturally combined with each other to yield the specifications for a broad range of categories, such as determiners, mass nouns, count nouns, maximal and submaximal nominal projections. This inventory of specifications allowed us to describe certain phenomena concerning the distribution of determiners without any devices of dubious theoretical status, such as empty categories, unary rules, or disjunctions.

However, it also turned out that our definition of submaximal projections was inappropriate to provide the distinctions necessary for a proper treatment of attributive adjectives. In section 4.5 we therefore first discussed some potential alternatives to the treatments of modifiers as adjuncts, in particular the approach pursued in the categorial grammar paradigm of treating modifiers as functors. We came to the conclusion that such an approach does not easily carry over to HPSG and that the insights and the advantages gained from a distinction between the notions of selector or functor should not be lightheadedly abandoned. We then proceeded to show that our approach to functional heads and in particular the split of head features and head domains allow us to solve the problem on a purely lexical basis, simply by introducing some additional `MINOR` feature and again without having to take recourse to empty categories or unary projections.

Although section 4.6 is somewhat marginal to the question of functional heads proper, it nevertheless addresses a type of construction that we had identified as a problem area for pure NP- and DP-approaches, viz., the second type of nominal constructions with a missing head, those without an overt substantive head noun. We discussed a broad range of different solutions, ranging from nominalization to coordination reduction, and in the end came to the conclusion that various aspects, amongst others the anaphoric nature of this phenomenon, point to the assumption of an empty nominal head. However, if such an empty head noun is postulated, the appropriate description of its distribution becomes a non-trivial task, since determiners, adjectives and nouns now are all optional or non-overt constituents, which nevertheless must not all be simultaneously absent in a nominal phrase. Again, we discussed a series of solutions and the technical problems involved and decided in favour of an approach which builds on a new type of feature percolation, which is different from the head feature projection in that it refers to properties of the first daughter or the left periphery in a phrase. In anticipation of the following section on declension, we showed that the assumption of such an edge feature, marking the presence or absence of an empty head noun, allows us to describe certain phenomena arising from a complementary distribution of certain determiner forms.

In section 4.7, we address one of the most diffuse and unwieldy problems of German, viz., the phenomena related to the different declension classes in the area of nominal categories. Our approach to this phenomenon differs from others in that it attempts to classify forms not according to their stem classes and syntactic function, but rather with respect to their morpho-syntactic distribution and inflectional form. We gain substantial advantages from the fact that we can treat the relationship between determiners and their nominal complements as a type of government, where declension as one of the `MINOR` properties can be different or identical to the one on the determiner as a function of its inflectional form. As we just showed in our comparison with the original approach of [Pollard and Sag 1994], the relation of the determiner governing the nominal complement is considerably more adequate and straightforward than the assumption that the adjective selects the appropriate determiner.

This approach to the treatment of declension phenomena is extended and applied to numerals in section 4.8. The purpose of this section is twofold. On the one hand we wanted to explore the generality of our assumptions on inflected and uninflected determiners. In this respect, we had to admit that there are certain limits and that other ways of expressing the generalizations may capture the constructions involving numerals more adequately. However, the second and potentially even more important result of this section is that a slight relaxation of our definition of functional heads will allow us to capture the occurrence of more than a single functional category with one substantive head. If we do not assume that every functional head must encode functional completeness, we can allow the introduction of even more than a single functional head domain within a phrase comprising all the advantages and flexibility required for the treatment of complex government and word order phenomena, as well as mutual dependencies between the co-occurrence of functional heads.

In this section we finally compared our approach with the original approach by [Pollard and Sag 1994] and its revised version of chapter 9. As it can be expected our approach fares extremely well in comparison to the original version in terms of empirical adequacy, simplicity and elegance of description. The differences are less drastic with respect to the revised version in terms of observational adequacy and weak equivalence. However, even for the revised approach we can still claim that we can achieve the same effects with less descriptive devices, specifically that we can dispense with two phrase structure schemata and replace them by lexical definitions and standard principles. The proof of this claim has been given in some part in this chapter and will be completed in the next chapter, where we will apply our theory of functional heads to the verbal domain.

Chapter 5

Functional Heads in the Verbal Domain

5.1 Introduction

In this chapter, we discuss the extension of our concept of functional heads to the clausal domain. In particular, we investigate the relation between complementizers and the position of the finite verb, for which we make a distinction between V-final clauses and V-initial clauses with the latter subsuming both V-second and V-first clauses. In the course of this discussion, we will also address the question of clause structures, arguing that the kind of analysis chosen for the so-called *Mittelfeld* will crucially determine the possible choice of solutions for analyzing finite verb position.

In this introductory section of the chapter, we introduce some very basic descriptive generalizations about German clause structure based on a topological model which distinguishes several positions or fields in clauses. In section 5.2, we compare different approaches to the analysis of V-initial structures and show that each of these analyses has to compromise in some respect. In section 5.3, we introduce our own analysis based on the notion of functional heads outlined in the previous chapter. In the final section 5.4, we address some open problems to our approach and introduce an alternative treatment of relative clauses.

German Clause Structure and Topological Fields

Many traditional descriptive analyses of German clause structure are based on a model of so-called topological fields. This model attempts to describe the structure of clauses (and sentences) by segmenting them into structural fields or positions, which are not or only loosely bound to specific grammatical relations.[1] In this model three different fields for so-called *Satzglieder*, i.e., grammatical functions other than the head, are assumed, which are bounded by two delimiter positions, the so-called *Satzklammern*:

Vorfeld	*Linke Satzklammer*	*Mittelfeld*	*Rechte Satzklammer*	*Nachfeld*

The *Vorfeld* (or initial field) in this model is claimed to be a position which can be filled by one constituent, be it a complement or an adjunct. This assumption has been frequently challenged on the basis of data which show that this position can be also filled by more than a single constituent (in particular in the case of adjuncts) but also by a partial constituent (as in the case of partial VP-fronting).

[1]Cf. for example [Drach 1939] as one of the earliest proposals along these lines. [Höhle 1983] gives an overview to the historical development of these ideas beginning in the last century.

However, by and large this assumption appears to hold and we will ignore these special cases below.

The *Mittelfeld* (or middle field) is supposed to comprise an arbitrary sequence of complements and adjuncts. The fact that these complements and adjuncts can basically come in any order, so that even arguments of different functors can be mixed, appears to be one of the main motivations for postulating this field rather than assuming a fixed configurational constituent structure with a contoured VP.

The *Nachfeld* finally is assumed to accommodate all right-extraposed constituents. These constituents are typically verbal projections, i.e., finite or non-finite complement or adjunct clauses. However, the *Nachfeld* is also a position which under certain circumstances is taken by non-clausal adjuncts or even complements.

In general, the motivation behind postulating these fields is to describe German as a non-configurational language which does not have fixed structural positions relative to which grammatical functions such as subject, object, indirect object, adjunct etc. could be defined. Thus, any of the topological positions mentioned, with the possible exception of the *Nachfeld*, can be taken by constituents with any of these grammatical functions. Of course, there have also been several attempts to reduce German clause structure to a configurational model. However, practically all of these approaches either fail to capture the full range of word order varieties or they have to take recourse to the stipulation of traces or transformations to construct a configurational structure in which structural positions unambiguously correspond to grammatical functions.

Asymmetries between Fields

While the topological model has its merits in capturing some descriptive generalizations, it would be highly fallacious to take the model as a basis for a formal description in terms of constituent structure. In particular, this model appears to suggest some structural symmetries for which there is quite obviously no evidence whatsoever. For example, while the distinction between a *Vorfeld* and a *Mittelfeld* is justified by the fact that the *Vorfeld* can be filled by constituents which cannot occur in the *Mittelfeld*, the implied symmetry between the *Vorfeld* and the *Nachfeld* is hardly motivated, nor is the relation between *Mittelfeld* and *Nachfeld* comparable to the one between *Mittelfeld* and *Vorfeld*. Similarly, the symmetry between the *Linke* and *Rechte Satzklammer* is only a nominal one which again does not require a direct reflex in a formal account of German clause structure.

The evidence that the *Vorfeld* is to be distinguished from the *Mittelfeld* is manifold.[2] To mention just a few, the *Vorfeld* is a position which is not limited to grammatical functions related to the local head domains defined be the clausal head, but

[2] Cf. among others [Reis 1980].

can be taken by constituents which stand in a genuine non-local dependency to an arbitrarily deeply embedded head. Constituents in the *Mittelfeld* on the other hand have to be related to verbal heads which appear in the clause-final verb cluster. Even if this constraint could be relaxed to allow extraction out of extraposed non-finite verbal complements (291), the positions in the *Mittelfeld* cannot be taken by a phrase which is extracted from a finite clause (292).

(291) weil es dieses Buch versucht hat, mehrmals zu lesen.

(292) Diesen Mann hat er behauptet, habe Maria gesagt, habe sie getroffen.

(293) * Er hat diesen Mann behauptet, habe Maria gesagt, habe sie getroffen.

Also, the *Vorfeld* position is a topological position which is essential in defining the notion of V-second clauses. To establish this construction, the *Vorfeld* position has to be filled by some grammatical function. If neither complement nor adjunct functions occur in this position, it can be taken by an expletive (294) whose occurence is excluded in the *Mittelfeld* (295).

(294) Es hat sich ein Unfall ereignet.

(295) * Ein Unfall hat es sich ereignet.

On the other hand, constituents in the *Vorfeld* can undergo some ellipsis (unless they function as expletive.), which is clearly not the case for *Mittelfeld* constituents.

(296) Was hast du gestern gemacht?
Habe ein Buch gelesen.

(297) Was hast du gestern gemacht?
Ein Buch habe gelesen.

In short, the kind of constituents, grammatical functions, and processes we find related to the *Vorfeld* is apparently a superset of what appears to be admissible in the *Mittelfeld*.[3]

Regarding the *Nachfeld*, the relation to the *Mittelfeld* looks more like the reverse subset relation. Although subordinate, adjunct and relative clauses strongly tend to get extraposed to the right, most of the constructions we find in the *Nachfeld* are also possible in the *Mittelfeld*, but not the other way round. Quite trivially this means that the *Nachfeld* constituents are a genuine subset of the *Vorfeld*

[3] An exception could be the occurence of the pronoun *es* which in its function as an accusative object can occur initial to the *Mittelfeld*, but not in the *Vorfeld*. However, there may be phonological explanations to this phenomenon.

constructions. In short, the *Vorfeld* and the *Nachfeld* are by no means symmetrical fields but rather two structural positions underlying quite different constraints.[4]

Parenthetical Dependencies

Besides the fields for the non-verbal constituents, the topological model also involves the relationship between the left and right delimiters of the *Mittelfeld*, the so-called *Linke* and *Rechte Satzklammer*. Again, it should be clear that the nominal symmetry between the *Satzklammern* is mainly an auxiliary concept, which is above all motivated by some mutual dependencies between specific occupations of clause-initial and clause-final positions. Essentially, the structural instances to be found can be reduced to the patterns (298).

(298)

Pattern	VF	LSK	MF	RSK
(a)	$XP_{(WH)}$	V_{fin}	XP *	V_{inf}*
(b)		V_{fin}	XP *	V_{inf}*
(c)		CPL	XP *	V_{inf}* V_{fin}
(d)	XP_{WH}		XP *	V_{inf}* V_{fin}

Some typical instances of these patterns are declarative or WH-interrogative main clauses (a), Y/N-interrogative, and imperative main clauses (b), subordinate clauses with complementizers functioning as complements or adjuncts (c), and subordinate WH-interrogative and relative clauses (d).

(299)

Pattern	VF	LSK	MF	RSK
(a)	Gestern	hat	er ihn	gesehen
(a)	Wann	hat	er ihn	gesehen
(b)		Hat	er ihn	gesehen
(b)		Sieh	ihn	an
(c)		daß / ob	er ihn	gesehen hat
(c)		als / weil	er ihn	gesehen hat
(d)	wer / der		ihn	gesehen hat

The dependencies holding between the occupants of the left and right positions are quite varying, both in terms of the categories involved as well as the obligatoriness of their occurence. For the patterns (a)/(b) a relationship between field delimiters can only be observed in the form of a dependency between the clause-initial finite verb and (potential) clause-final non-finite verbs governed by it, but also between

[4]Since the question of right-extraposition is not only a very complex problem but also somewhat marginal to the issues we want to raise, we will more or less ignore the *Nachfeld* in the following.

a finite verb and its separable prefix. Since the right-hand side does not necessarily have to be realized, this relationship cannot always be based on pure observation.

The dependency between the clause-initial complementizer and a clause final finite verb (c) is the most consistent one, since the existence of a complementizer is always bound to the occurence of a finite verb in a clause-final position.[5] For the pattern (d), it is descriptively and theoretically[6], quite unclear whether the clause-initial WH-phrases should be better assigned to the *Vorfeld* or topic position, as we did here, or whether it should be considered an instance of the LSK. This analysis of this construction is additionally complicated by the fact that here the left-hand side boundary of the *Mittelfeld* is marked by a constituent which need not be a lexical item, but can also be a complex phrase. Thus, the dependency between the *Linke* and *Rechte Satzklammer* in this case cannot be described as a lexical relationship or at least a relationship between two lexical items. In short, it appears that the notion of a parenthetical structure is manifested most clearly by constructions involving a complementizer, while it seems to have limited descriptive value for all other constructions, at first sight.

Nevertheless, the topological model implicitly also captures two very crucial generalization about German clause structure. The first one is observable in the relationship between the patterns (a)/(b) on the one hand and (c) on the other and concerns the complementary distribution between complementizers and finite verbs in the clause-initial position. The second one concerns the fact that all clauses of which a *Satzklammer* could be predicated in some way or other have to be finite clauses. Quite different from English, the observed parenthetical patterns are hardly as clearly manifested in non-finite constructions, but always have some alternative explanations. No non-finite construction ever has a verb in an initial position, and the few candidates of potential non-finite complementizers (i.e., *um* and *anstatt*) can easily also be interpreted as prepositions.[7]

In the following sections we will concentrate on these issues, viz., the complementary distribution of complementizers and finite verbs in clause initial position, as well as the inherent relationship between the notion of a *Satzklammer* and the occurence of a finite verb. We will attempt to capture the relevant phenomena on the basis of an integrated analysis building on our notion of functional heads.

---

[5] There are some minor (colloquial) exceptions to this generalizations, as for example the Bavarian V-second construction with the prefixed adverbial *weil*,

(i) weil er hat das nicht gewußt

[6] Cf. among others [Reis 1985], [Scherpenisse 1985] and [Haider 1987].

[7] Cases involving interrogative or relative pronouns, as for example (i) or the pied piping constructions (ii), are either quite marginal or also have alternative explanations:

(i) Wohin sich wenden in seiner Not!?

(ii) ... ein Buch, das zu lesen man nicht missen sollte.

5.2 Approaches to V-Initial Structures

Our main concern in this section will be to explore the relationship between finite verb position and the analysis and structure of the *Mittelfeld* in German. For this purpose, we will discuss different approaches to the analysis of the initial and final position of the finite verb in German, as they are exemplified in (300) and (301). As mentioned, we consider V-second clauses as a structural variant of V-initial structures which is irrelevant for the issue at hand.

(300) [$_H$ Bringt] [$_C$ er] [$_A$ morgen] [$_C$ dem Kunden] [$_A$ pünktlich]
[$_C$ die Fracht]

(301) daß [$_C$ er] [$_A$ morgen] [$_C$ dem Kunden] [$_A$ pünktlich]
[$_C$ die Fracht] [$_H$ bringt]

The different analyses assigned to these structures have often been grouped into *flat* and *contoured* analyses.[8] However, it appears that the space of possible accounts in HPSG and related frameworks, such as Categorial Grammar, can be also broken down in terms of three different factors or parameters which are as follows.

(a) The standard mechanisms and principles for subcategorization and adjunction are maintained.

(b) V-initial and V-final clauses are assigned a uniform structural representation with respect to the structure of the *Mittelfeld*.

(c) Finite verbs in clause-initial position are analyzed as heads *in situ*, i.e., they are assigned the grammatical function of a `HEAD-DTR`.

Although it may in principle be desirable to maintain all of these basic tenets, we will argue that only two of the basic assumptions can be maintained simultaneously. Maintaining all three assumptions will require the introduction of additional devices leading to a modification and extension of fundamental principles and generalizations of the framework.

In our own approach, described in the next section, we will show that for those approaches which attempt to maintain all three assumptions by introducing additional mechanisms, an equivalent solution can be found which abandons the assumption (c) but requires less additional devices. We motivate this approach not only by the fact that abandoning (c) has the least global effect on the theory

[8]It should be clear that the latter term is not equivalent to the assumption of straightforward VP, but rather subsumes all clause structures in which not all complements and adjuncts are sisters to each other.

and possible descriptions of the clause structure, but also by integrating it with an analysis of V-final clauses introduced by complementizers. Our approach will allow to describe the *Mittelfeld* homogeneously as a binary right-branching structure without extensive modifications of general principles of the theory. The key to the analysis is the postulation of a clause-final empty node as a functional head which takes the finite verb as a complement.

Among the existing and possible analyses, we first discuss the quite prominent approach taken by [Pollard 1991] and others, which maintains the parameters (b) and (c) by postulating a flat structure encompassing the finite verb and its complements and adjuncts. We will argue that this approach has to abandon assumption (a) and not only complicates the treatment of adjunction, but also commits one to the assumption of a phrase structure for which there is little linguistic motivation.

As a variant of the rejection of (a) we discuss the possibility of a uniform binary right-branching structure which involves a modification of the SCP and which has a technically very unpleasing analysis as a consequence.

Thirdly, we will discuss the effects of maintaining the assumptions (a) and (c) on the second parameter, above all the implications of a non-uniform phrase structural representation for V-initial and V-final clauses.

Finally, several different solutions suggested by [Kiss and Wesche 1991] building on [Borsley 1989], [Oliva 1992] and [Frank 1994] attempt to maintain all three assumptions at the cost of introducing additional devices. These approaches also have in common that they are all based on the assumption of a binary right branching clause structure.

5.2.1 Flat Clause Structures

Within HPSG, the most prominent approach building on a flat clause structure is presumably the one suggested by [Pollard 1991] who thus attempts to preserve the assumption of a uniform representation of the MF for V-initial and V-final clauses as well as the assumption that the finite verb functions as a head irrespective of its position. Since Pollard also does not want to abandon the assumption that all languages should be analysable by a small set of universal phrase structure schemata, he suggests to treat these flat clause structures in German as a more general case of the phrase structure schema which was originally employed for the analysis of auxiliary inversion in English. The finite verb is accordingly treated as the head of the clause which in V-initial and V-final clauses appears as a left- or rightmost sister to the MF constituents.

The relevant phrase structure schema invoked for this approach is defined in [Pollard and Sag 1994][40] as follows:

(Schema 3) a saturated ([SUBCAT $\langle\ \rangle$]) phrase with DTRS value of sort *head-comp-struc* in which the HEAD-DTR value is a lexical sign.

For English this schema is restricted to apply only to phrases headed by an auxiliary or modal verb. The finite forms of these verbs are underspecified with respect to a topological HEAD feature INV marking inversion, while all non-finite forms and all main verbs are marked as non-inverted, so that they are blocked from occurring as heads of these constructions.

The parametrization of this schema for German involves a neutralization of the auxiliary – main verb distinction so that all finite verbs have an underspecified INV feature and the schema can be applied to all constituents headed by a finite verb. In fact, for finite constructions in German only this one schema is envisaged, licencing both V-initial and V-final clauses. All root sentences (or main clauses) are then constrained to be marked as [INV +]. Two LP rules have to specify that a verbal head marked as [INV +] has to precede all its complements, whereas a verbal head marked [INV −] has to follow them.

The structure for V-initial clauses will then be roughly as given in (302).

(302) [[$_H$ Schenkt] [$_C$ der Mann] [$_C$ dem Kind] [$_C$ ein Fahrrad]]

The two main advantages of this approach are that it not only provides a solution to the problem of finite verb position but also offers a comparatively simple basis for a treatment of the free or relaxed order of complements in the MF. Since the subject and all complements are bound simultaneously by the finite verb in the same schema, they can be freely permuted and can appear in any order.

What are then the advantages and disadvantages of this approach? As it seems, many arguments in this context build on the opposition between flat and contoured clause structures, and attempt to argue on a more or less empirical basis in favour of one or the other type of constituent structure. The most prominent of these issues are briefly sketched in the following.

Clause union phenomena are mainly those which seem to suggest that under certain conditions the main, modal and auxiliary verbs may form a verb cluster within the boundaries of a finite clause and that the complements of these verbs can be freely permuted.[9]

One technique to deal with these phenomena, which has almost become an HPSG standard, was developed by Hinrichs and Nakazawa[10]. The basic idea of this

[9] Cf. [Netter 1991] for a broader range of the phenomena and issues associated with clause union and verb raising.

[10] Cf. [Hinrichs and Nakazawa 1989] and [Hinrichs and Nakazawa 1994]

approach is that a matrix verb can (lexically) raise the subcategorization of an embedded (non-finite) verb by appending this subcategorization to its own. As a result the valence requirements of embedded non-finite verbs can be raised to the topmost level of the finite verb, so that all complements of a verb cluster can appear on a par with those subcategorized for by the finite verb itself. Although this approach was originally developed to show how auxiliary flip can be treated in a contoured VP analysis, Pollard shows that in principle it can also be exploited in an approach aiming at a flat clause structure.

However, although Pollard does not discuss the cases of auxiliary flip in great detail, the adaptation of the Hinrichs/Nakazawa approach to a flat analysis might come to its limits, if more than one auxiliary is flipped, as in (303).

(303) weil er es [hat_1 [wollen_2 [[lesen_4] lassen_3]]]

(304) *weil er es [[wollen_2 [[lesen_4] lassen_3]] hat_1]

The generalization underlying the difference in grammaticality between (303) and (304) is that the recursion within a verb cluster can start off by a right-branching and then turn into a left-branching structure, but it can never switch from a left-branching into a right-branching recursion (and back again), as it is the case in (304). One rather convincing approach to this phenomenon was given by [Johnson 1986] who argued that the difference in grammaticality between (303) and (304) should be described by a mixture of lexical and structural constraints propagated within the verb cluster.

Pollard's approach faces the following dilemma with respect to this problem: either he assumes that the finite verb is part of the verb cluster, which means that the head of schema 3 cannot be a purely lexical constituent anymore, or some structural relationship has to be established within schema 3 which transports information about the position of the finite verb and the branching direction of the verb cluster.

On the other hand, it is well possible that one should not overrate this issue too much, as long as no convincing solution to another problem of auxiliary flip is found. Since the Hinrichs/Nakazawa approach builds on some context-free recursion, the following construction defies an analysis in this as well as—to our knowledge—in all other existing approaches.

(305) weil er es lesen_3 hat_1 lassen_2

Partial VP (PVP) fronting is a phenomena which has often been taken to argue for a contoured clause structure (as well as against a fixed configurational structure involving a standard VP).[11] One of the main concerns of Pollard's paper is therefore to show that PVP fronting can be accommodated with a flat analysis in combination with SUBCAT raising.

To capture these constructions it is suggested that the head-complement schema 2 is modified such that it does not require the consumption of all non-subject complements on the `SUBCAT` (or `COMPS`) list anymore, but that it will also licence structures in which only a subset of the complements is bound. This means that for example a PVP constituent can be formed which does not contain all objects. From a fronting or topic position such a partially saturated constituent can then be bound to a trace in the *Mittelfeld* and the remaining unsaturated subcategorization requirements can be raised by the standard techniques.

However, while the relaxation of the constraints of schema 2 allow to capture the quite marginal PVP fronting cases, it also leads to a massive amount of spurious ambiguities in the vast majority of other clauses, since any complement can now be consumed at any arbitrary level of embedding. A solution to this problem is suggested by [Nerbonne 1994b] who shows that the spurious ambiguities arise from the assumption of a trace and that they can be neutralized if the fronting of partial VPs is treated by means of a lexical rule.

Nerbonne argues quite correctly that the only evidence for the assumption of a trace is the methodological assumption that if a dislocated substring is treated as a constituent in one place it must be treated as a constituent throughout the clause structure, in particular in its "base position". A lexical approach to non-local dependencies on the other hand is not bound to such an assumption, i.e., even if a PVP is treated as a constituent in the fronted position, this does not necessarily imply that this substring must also be able to occur as a constituent in the *Mittelfeld*, but rather the information associated with that partial constituent can be dispersed over different levels of embedding.[12]

More importantly, however, as soon as one accepts this move of disassociating a fronted PVP constituent from the phrase structure in the *Mittelfeld*, PVP fronting cannot be adduced as an argument in favour or against either a flat or a contoured *Mittelfeld* structure anymore.[13]

[11] Cf. among others [Nerbonne 1986] and [Uszkoreit 1986b]

[12] Note by the way that this assumption also need not hold in non-lexical approaches to non-local dependencies, as long as they dispense with trace mechanisms. This same basic assumptions also underlies the LFG functional uncertainty approach to word order phenomena and constructions involving partial constituents in German, as it was outlined in [Netter 1986], [Netter 1987] and [Netter 1988].

[13] [Nerbonne 1994b][110] himself explicitly emphasizes that he uses Pollard's flat analysis as a basis for his approach because it makes the least assumptions about constituent structure, but

Coordination data have been used repeatedly to argue for a contoured analysis, which is quite natural because a more structured constituency will always offer more options for treating a coordination phenomenon as constituent coordination than an entirely flat structure. On the other hand, these phenomena have been also taken to argue against a more contoured analysis, since clearly not all data can be explained consistently on that basis. The problem is simply that coordination encompasses such a broad range of different phenomena typically interacting with each other that it is hardly possible to find a one and all explanation for the entire range of data and that it is not obvious whether a subset of the data can be isolated for which a consistent analysis can be found. Moreover, even if this were the case, the chances that a weakly equivalent analysis can be found are not negligible. Thus, we will only briefly review some of the most relevant data below, but we will be rather cautious to interpret the data in one direction or the other.

As [Nerbonne 1994b] points out there are a number of coordination constructions which under a flat analysis have to be treated as non-constituent coordination. The emphasized constituents in (306) for example are sisters to the subject and the verb, but do not form a constituent on their own.

(306) Gestern [[erzählte] [Hans] [*seinem Sohn*] [*ein Märchen*]
und [*seiner Tochter*] [*eine Abenteuergeschichte*].

Nerbonne rejects these data as an argument for constituency on the basis of the claim that non-constituent coordination is very common in German. This is indeed the case in right-node raising or right periphery deletion constructions such as (307) where the coordinated strings may even cut across word boundaries:

(307) Gestern hat [*Hans den Auto-*] und [*Peter den Teppich-*Händler verschaukelt.]

However, if sentences such as (306) were indeed interpretable as the mirror-image of sentence (307), i.e., if they were a case of "left-node-raising" or "left periphery deletion" one would also expect that genuine non-constituents could also be coordinated on the right-hand side of a sentence. Yet, a sentence corresponding to (307) such as (308) appears to be quite out of question:

(308) * Gestern hat [Hans den Auto-*Mechaniker bezahlt*] und [*-Händler verschaukelt.*]

The closest that one gets to non-constituent coordination on the right-hand side are sentences involving a cut through a prepositional phrase [Nerbonne 1994b][121]:

(309) Er fährt [mit *Tom nach Paris*] und [*Max nach London.*]

However, as it seems all other attempts to cut across noun phrase internal structures appear to become increasingly less acceptable:

---

that his solution is neither committed nor confined to this analysis.

(310) *Er traf [Pauls *Mutter in der Disco*] und [*Vater im Supermarkt.*]

(311) *Er mag am liebsten [Brot *mit Nutella zum Frühstück*] und [*ohne Aufstrich zum Abendessen .*]

As already mentioned, we do not intend to derive an argument for a contoured analysis from these data and we would only commit ourselves to the claim that clear non-constituent coordination cannot be invoked to weaken the force of the argument derived from sentences which could be covered by a constituent coordination analysis under the assumption of a contoured clause structure.

Scope of quantifiers, adverbials, and binding is another phenomenon which has repeatedly been used as evidence for a contoured analysis. The argument is typically based on the assumption that c-command, as the basis of scope relations in a right-branching structure, will yield different and less constrained results in a flat analysis. Again, it appears that there is a large set of cases for which this assumption can yield the correct results. On the other hand, there are also some data which appear to contradict this assumption by having an alternative wide scope reading of a more deeply embedded constituent over an element to its left. Some of these examples even permit only a wide scope reading due to semantic incompatibility.

Both [Nerbonne 1994b][122ff] and [Kasper 1994][44ff] discuss such cases in great detail, arguing against the conclusiveness of the evidence. While their argument to a large degree is quite valid, we at least want to point out that in most of the cases presented as counterarguments there appears to be a clear difference in markedness. We pick out only two examples from the data presented. In the case of (312) both sentences must have the same scope interpretation. However the word order of the second sentence (b) is clearly more marked than the one of (a).

(312) (a) Peter hat täglich 3 Stunden lang trainiert
(b) Peter hat 3 Stunden lang täglich trainiert

Similarly, in the case of (313) the scope of the adverbials is presumably ambiguous, however, the preferred reading appears to be the one which follows the order of elements on the surface.

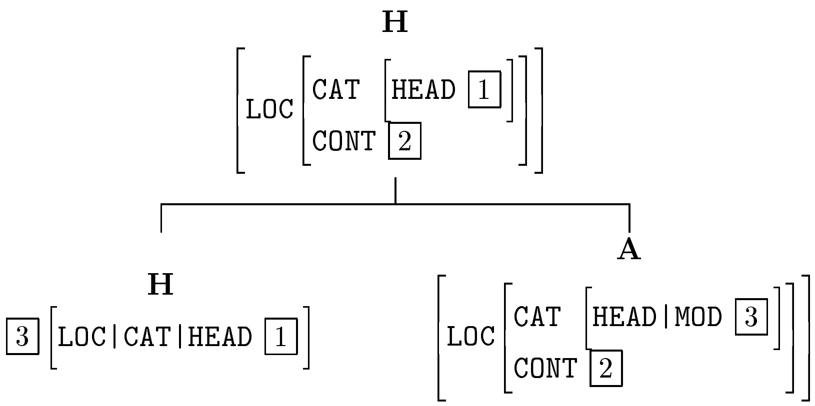
(313) Das Baby heulte ab und zu eine Stunde lang.

Thus, while it would be untenable to claim that scope directly correlates with surface order or c-command, there at least seems to be a strong tendency for this correlation. Whether this is more appropriately captured by pure linear order (in a flat structure) or whether c-command in a contoured structure, may well be a purely technical issue which depends on the way how scope is read off from and represented in a structure.

Adjunction is finally an issue which we considered as decisive enough to explore an alternative to Pollard's flat analysis. While most of the phenomena discussed above do not provide any conclusive empirical evidence, the treatment of adjunction in this approach appears to become so complicated as to prompt the search for an adequate contoured analysis. Specifically, it appears that if one wants to maintain a uniform flat structure for the MF as well as the assumption of the finite verb as a head *in situ* in both V-initial and V-final clauses, one may be able to preserve the Subcategorization or Valence Principle, but not the treatment of adjunction as it is proposed in [Pollard and Sag 1994][55ff].

We briefly recapitulate the essentials of the treatment of adjunction in standard HPSG for ease of reference. Adjuncts in HPSG are characterized as non-head daughters which select a head through an attribute MOD whose value is identified with the SYNSEM value of the HEAD-DTR. In the Semantics Principle (SemP) the flow of semantic information for head-complement structures and head-adjunct-structures, repeated in (314), is defined such that in the latter case the CONTENT value of the mother node is not identified with the one of the HEAD-DTR but rather with the CONTENT value of the ADJ-DTR. Thus, while the HEAD-DTR continues to provide the syntactic head, the ADJ-DTR functions as the semantic head.

(314) Phrase Structure Schema for Head Adjunct Structures



In this approach, it is essential that head-adjunct structures are binary branching structures. As a consequence, sentences such as (315) where several adjuncts are interspersed with complements, present a serious problem for the flat *Mittelfeld* structure suggested by [Pollard 1991].

(315) [$_H$ Bringt] [$_C$ er] [$_A$ morgen] [$_C$ dem Kunden] [$_A$ pünktlich] [$_C$ die Fracht]

To accommodate the flat analysis with such a construction, it will be necessary to modify the approach such that in the flat head-complement structure the verb can combine not only with all the complements but also with all the adjuncts at the same time. For a start, this means that the rule schema 3 has to be extended so that it includes not only the `HEAD-DTR` and an arbitrary number of complements, but also an arbitrary number of adjunct daughters, i.e., the value of the attribute `ADJ-DTR` will have to be a list or set of feature structures rather than a simple feature structure.

Since each element of this set of adjuncts will impose selectional restrictions on the head, universal quantification is necessary in order to distribute the head daughter's information over all the elements. Also, quite obviously the SemP as it was specified above would cause conflicts, since now a head–complement configuration, in which the syntactic head coincides with the semantic head, is merged with a head–adjunct configuration, in which the adjunct functions as the semantic head.

This problem is the central concern of [Kasper 1994] who suggests to split the selectional constraints specified in the `MOD` feature of the adjunct into two parts. Through one feature the syntax of the verbal head is selected. The second feature is responsible for semantic selection and is compatible with the possibility that an adjunct may not only incorporate into its semantics the semantics of the lexical head but also by means of recursive selection the semantics of another adjunct. However, the formulation of both types of selections in a radically revised Adjunct Principle must build on universal quantification over the set of adjuncts in a rule schemata, i.e., the application of functional or relational constraints is inevitable.

Yet, even if the solution suggested by [Kasper 1994] compensates one of the main deficiencies of Pollard's approach, the issue nevertheless gives rise to the question whether a radically flat structure does not render the notion of phrase structure or constituency on a clausal level almost obsolete. All more or less fine-grained distinctions and grammatical relations implied by the other rule schemata are almost entirely neutralized and lumped together into one single schema, such that it appears almost feasible to eliminate all other schemata altogether.

It should be kept in mind that rule schema 3 was originally designed as a very specific schema constrained to auxiliary inversion in English, which now ends up as a maximally general schema into which almost all other schemata, accounting for head–subject, head–complement and head–adjunct relations, are merged. The question also arises why this "generality" should be limited to German and does not seem to have a reflex in the grammars of other languages.

In our own approach, which we already outlined in section 3.4, we will therefore pursue a different direction. Rather than merging all grammatical relations into one quite unconstrained schema, we will modularize and break down these relations into minimal binary branching units, shifting the necessary modifications into the lexicon as well as maintaining all the standard principles as they are.

5.2.2 Inversion of the Subcategorization Principle

As already [Kasper 1994] points out, the simplest solution to solve the problem of adjunction in combination with complementation would be a binary branching structure in which the verbal head is combined with only one complement or adjunct at a time.

While the head-adjunct rule could be maintained as it is given above, basically only one head-complement schema would be required (316). For expository reasons we ignore the finer distinctions outlined in section 3.4 above and simply assume that the Subcategorization Principle (SCP) is reduced to a simple function **cons** concatenating the `SYNSEM` value of the `COMP-DTR` with the `SUBCAT` list on the mother.

(316) Phrase Structure Schema for a Binary Head-Complement Structure

$$\overset{\mathbf{H}}{\left[\texttt{LOC}\left[\begin{array}{ll}\texttt{CAT} & \left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{SUBCAT} & \boxed{2}\end{array}\right]\\ \texttt{CONT} & \boxed{3}\end{array}\right]\right]}$$

$$\overset{\mathbf{H}}{\left[\texttt{LOC}\left[\begin{array}{ll}\texttt{CAT} & \left[\begin{array}{ll}\texttt{HEAD} & \boxed{1}\\ \texttt{SUBCAT} & cons(\boxed{4},\boxed{2})\end{array}\right]\\ \texttt{CONT} & \boxed{3}\end{array}\right]\right]} \qquad \overset{\mathbf{C}}{\boxed{4}}$$

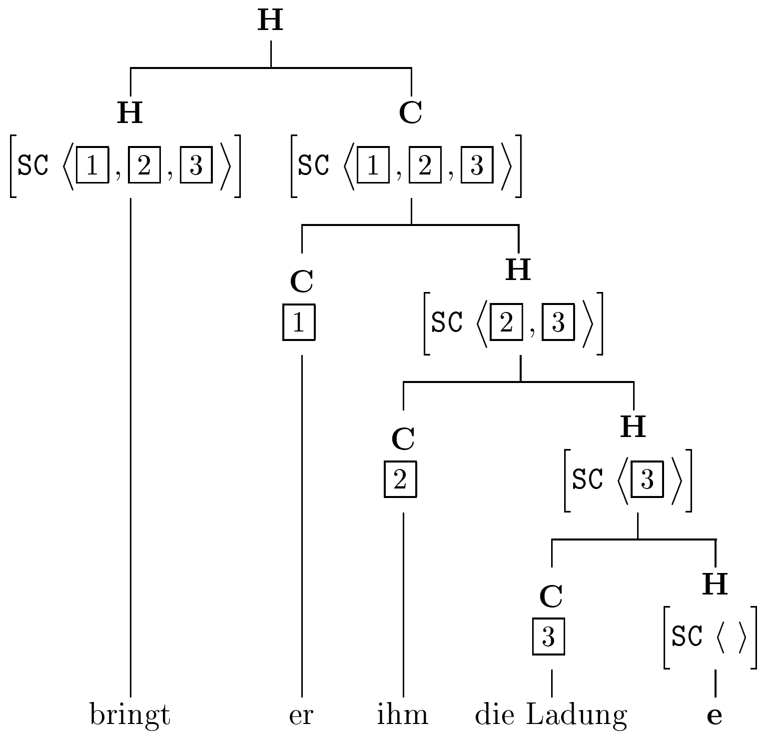
The advantage of such a binary derivation is quite obviously that complements can be freely mixed with any number of adjuncts, while the identity of the syntactic or semantic head is always well-defined at any given stage of a derivation.

However, if one wants to maintain in addition that both clause-initial and clause-final finite verbs are analyzed as heads *in situ* and that the MF is assigned a uniform structure in V-initial and V-final clauses, the SCP would have to be modified. This follows from the fact that a subcategorization list according to the standard SCP always has to be reduced 'inside out'. Since the head introduces the subcategorization list which is then successively reduced until it has become an empty list at the root of the head domain, the position of the head will automatically determine the branching factor. Heads in a left peripheral position can only be embedded in left-recursive structures right-peripheral heads only in right-recursive binary structures.

Since a uniform, left- or right-branching structure would run against this basic generalization in one way or another, the SCP has to be inverted locally for one of the two constructions, so that it works exactly the opposite to standard derivations. In such a definition, the `SUBCAT` value of the mother node would have to be the sum (rather than the difference) or concatenation of the `SUBCAT` lists of the `HEAD-DTR` and the `COMP-DTR`.

Such an analysis could be implemented, for example, by making the finite verb select the *Mittelfeld* as a constituent and to copy its `SUBCAT` frame to the root of this constituent. Within this phrase the SCP would have to be inverted locally to reduce the complementation list in a top-down manner. At the bottom of the structure an empty head-node would guarantee the termination of the recursive reduction.

(317)



For V-final clauses the standard SCP could be applied as before. However, the top-most rule combining the verb with the MF could also be interpreted as a dominance rule such that the verb-final and verb-initial clauses can be derived through two different LP statements. Such an analysis could be seen as a variant of the approach of Pollard described above, differing only with respect to the representation of non-verbal complements and adjuncts as a structured constituent rather than a flat list. As far as we can see, the treatment of other phenomena, such as verb-raising, partial VP-topicalization etc. could be accommodated within this analysis in an analogous way.

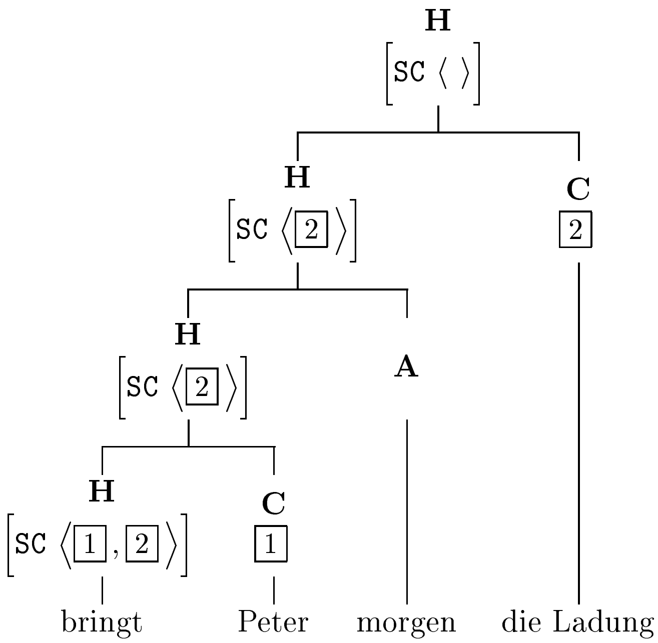
The main problem with such an approach is that the application of the different variants of the SCP would have to be triggered either by duplicating structural rules or lexical entries. In any case, it would lead to a proliferation of the number of purely diacritical features whose single purpose is to encode the possible combinatorics of rules and principles.

Note, however, that the basic idea of such an approach is not too farfetched, even if it may be hard to realize in HPSG. [Hepple 1990], for example, proposes an analysis in categorial grammar which goes into a very similar direction. Hepple suggests to apply type raising to the verbal head as well as to the complements, such that all complements can be combined with each other to form a constituent *before* they combine with the head.

5.2.3 Non-uniform Structural Representations

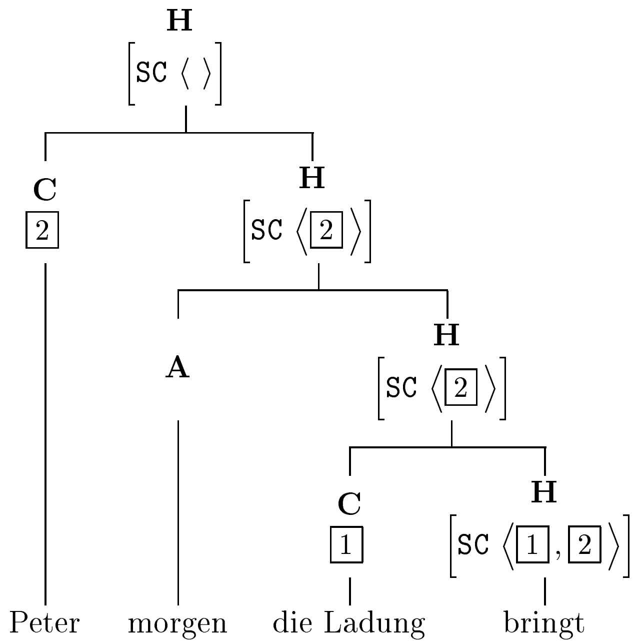
One solution in the space of possible options is to maintain the assumption of the finite verb as a head throughout as well as the standard principles and rules, which strongly suggests a binary branching structure, but to give up on a uniform representation for V-initial and V-final clauses. As just mentioned, this assumption will make the standard SCP enforce two different directions of branching for V-initial and V-final structures, namely left-recursive (318) and right-recursive respectively (319).

(318)



The advantages would be obviously that under these assumptions no extra mechanisms and no changes to the theoretical foundations of HPSG would be required.

(319)



However, the assumption of a left-branching structures as a mirror image of right-branching structures also has the effect that quite a number of relations typically associated with the constituent structure would have inverse realizations in the two types of structures.

To begin, standard c-command relations obviously will be inverted for different head positions. If c-command should play a role in determining the scope of quantifiers or adjuncts, we end up in a situation where different scoping relations will be associated with the same sequence of complements and adjuncts, depending on whether the verbal head appears in a final or initial position.

Similarly, the simplified implementation of the SCP above, which exploits the order of elements on the `SUBCAT` list for determining the order of complements, will now give us reversed orderings of complements for a left- and a right-branching structure. As a consequence, the order of elements on the `SUBCAT`-list for the two different verb positions has to be reversed, i.e., two different lexical entries may have to be assumed.

Of course, this may well be a spurious problem, which we mention mainly because an approach along these lines has been suggested in [Hoeksema 1985] for categorial grammar. The issue will not arise if the order of complements and adjuncts is determined independently by the specification of LP rules which are not restricted to sister constituents in a local tree of depth one.

However, even then this account enforces the status of a constituent for substrings for which there is hardly linguistic evidence that they will ever form a phrase in other circumstances. In particular, in a simple unmarked V–S–O structure we would have to claim that the verb first of all forms a constituent with its subject before it combines with an object. The only phenomena which would corroborate such an assumption are again bound to coordination. Specifically some cases of right-node raising could be treated as a constituent coordination phenomenon, at least for pure V-initial clauses.

(320) [[[Verkauft] Hans] Maria] oder [[[schenkt] Paul] Sophie] das Auto.

Unfortunately, even this advantage presumably will be gone in V-second clauses, since there it is quite likely that the topicalized phrase is added to the clause after the finite verb has been combined with its MF constituents. While the proposed structure can be accommodated in the case of a combination of right-node raising and (left-peripheral) topic deletion (321), it does not really lend itself very easily to right-node raising occuring on its own (322)—which only goes to show that coordination is hardly a reliable test for constituency in German.

(321) [Den Wagen [[[[verkauft] Hans] oder [[verschenkt] Paul]] an Maria]]

(322) Den Wagen [[verkauft] Hans] und das Fahrrad [[verschenkt] Paul] an Maria

Note, however, that the argument that this structure does not provide the constituency required for PVP fronting apparently does not hold anymore as soon as a trace-less solution is assumed. Yet, it is still the case that more or less the only advantage that we can obtain from a left-branching binary structure is a simplification of the treatment of adjuncts, whereas it does not increase in any respect the linguistic adequacy or elegance of the description. In comparison to the approach described in the preceding section, this analysis would be superior only in that it does not require a modification of the general rules and principles.

5.2.4 Right-Branching Binary Structures

To resolve the obvious conflict between a right-branching structure and a left-peripheral (verbal) head, several proposals have been made to transport the subcategorization frame from the V-initial position to the right periphery of the clause such that the standard SCP can be applied. In all of these cases some 'carrier feature' has been used or introduced to transport the information from the "external", clause-initial position of the head to some "internal", i.e., clause final position.

One proposal along this line was made by [Kiss and Wesche 1991] who apply an approach by [Borsley 1989] to German and who use some additional non-local mechanism. Another suggestion is given by [Oliva 1992] who exploits the Head Feature Principle (HFP) for this purpose. Finally, [Frank 1994] combines the approach by [Kiss and Wesche 1991] with our own approach to functional heads outlined above, but inadvertedly also incorporates some aspects of [Oliva 1992].

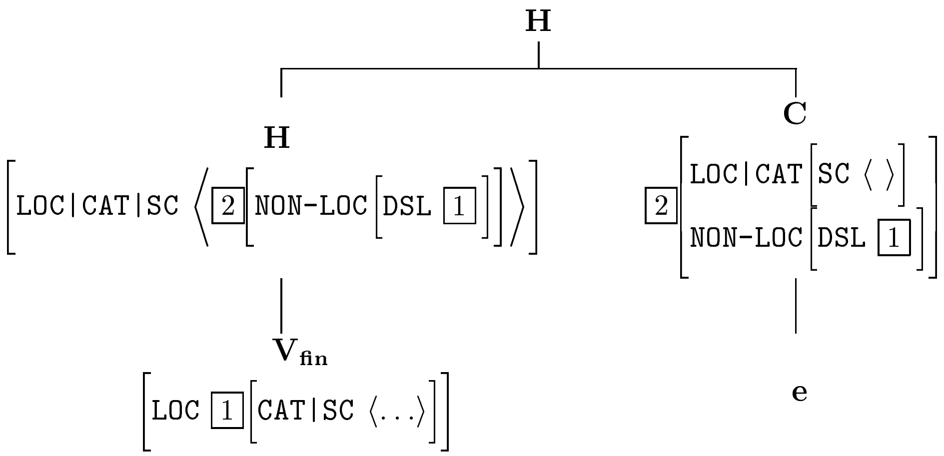
Kiss and Wesche propose to introduce a specific additional non-local feature 'Double Slash' `DSL` in analogy to the common non-local feature `SLASH`, which is used to percolate dislocated WH-Phrases in standard non-local dependencies. An essential characteristic of this approach is the application of a lexical rule which must be applied to the finite verb in a clause-initial position and which changes the subcategorization properties of this verb.

The result of this lexical rule is a verb which subcategorizes for a verbal or clausal complement which must be headed by an empty category. This verbal complement is assigned a `DSL` feature whose value is identical with the original `LOCAL` value of the verb. The `DSL` value, thus containing also the original subcategorization frame, is percolated to the bottom of the right-branching verbal projection which is anchored in the empty head. The empty head takes the subcategorization frame from the `DSL` feature and identifies it with its own, now selecting all the complements that the clause-initial verb required. Since the `SUBCAT` list is now to the right of the expected complements, it can be reduced bottom-up in a right-recursive structure by the standard SCP. The structure in (323) illustrates the most crucial part of the derivation, viz., the effects of the lexical rule, represented here as a unary domination on the finite verb, and the combination of the head with the verbal complement anchored in an empty terminal.

(323) Introduction of `DSL` (Including Lexical Rule)

At first sight our own solution presented below appears to be quite similar to this approach and we are quite willing to admit that the two solutions fall into the same class of analyses, in which the subcategorization information is transported from the left to the right periphery. However, at closer scrutiny the `DSL` approach suffers from several weaknesses, which are absent from our own analysis.

To begin, the analysis of [Kiss and Wesche 1991] requires the assumption of a lexical rule *in addition* to the assumption of an empty lexical head. This means that the V-initial/V-final variation has to be encoded by a combination of lexical and structural information, since finite verbs must be assigned different lexical properties depending on the structural position they hold. Thus, which of the two lexical variants of a finite verb is to be chosen can only be determined on a more global structural basis. Also, since both the finite verb in initial position



and the empty terminal in final position function as heads, two different LP rules for determining the order of head complements in verbal projections have to be envisaged, which have to be made sensitive to the three different types of finite verbal heads (basic, modified by lexical rule and empty).

The analysis employs a mechanism which was originally devised for non-local relations to express a dependency which is clearly limited to 'head domains', i.e., the local domain defined through the HFP. Genuine non-local dependencies in German, such as topicalization or dislocation of WH-elements, which are covered by the relations established through SLASH, are not constrained to the boundaries of maximal projections. More specifically, these non-local dependencies are typically relations which may hold across clause boundaries or at least across more than one head domain. The relation between a finite verb in initial position and a corresponding empty terminal node on the other hand is clearly bound to a domain which is even smaller than the domain of a clause, if one assumes that a clause may contain several (non-finite) verbal heads in addition. As a consequence the Non-Local Principle governing the distribution and flow of non-local information has to be artificially constrained for this particular case, so that the DSL feature can be confined to a head domain, i.e., that it can be percolated *only* along a path of HEAD-DTRs.

Last but not least, this approach fails to capture the complementary distribution of complementizers and V-initial finite verbs in any non-stipulative way. In order to prevent the combination of a complementizer with a V-initial clause, the lexical rule has to mark the V-initial verb by some diacritical feature, say INV+, which then can be referred to by the complementizer in its SPEC feature. Even if the validity of all other arguments we raised against this analysis were questioned, this is one crucial aspect in which we can claim a clear superiority of our own approach.

Oliva 1992 suggests another approach falling into the same class, which appears to be more satisfying at least with respect to the issue of locality of the dependency relation between the V-initial finite verb and the clause-final empty terminal. The central feature of this analysis is to exploit the HFP directly for establishing a relation between the subcategorization of a clause-initial finite verb and a clause-final empty node.

The structures [Oliva 1992] proposes are somewhat similar to the ones suggested by [Kiss and Wesche 1991], one crucial difference being that not the entire `LOCAL` value of the verb is transported but only its most relevant part, the `SUBCAT` value. As the carrier feature, Oliva uses a `HEAD` feature named `HEAD-SUBCAT`. Since this feature is only percolated by the Head Feature Principle (HFP) it is automatically confined to the respective head domain.

However, in some respect [Oliva 1992] dodges the issue of the grammatical relation between a clause-initial head and its sister constituent. He avoids this question by simply recoding right-branching binary structures into `FIRST`–`REST` lists.[14]

The `FIRST` elements in these structures are assigned some type, which encodes the grammatical function of the value of that structure, the `REST` feature is simple assigned the indiscriminate type *phrase*, presumably equivalent to a type *list* since it does not always have to be a phrasal constituent, of course. The only case foreseen in which the first element functions as a head is the V-initial construction, while in all other cases the head is to be sought in the second or `REST` constituent. This split of phrasal configurations is matched by a split of the HFP into two clauses, one, which is more specific, determining the (locally) head-initial structure (324)(a) and a second, more general one applicable to head-final configurations (324)(b).[15]

(324) (a) $\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{FIRST} & \mathit{head}\begin{bmatrix}\texttt{HEAD} & \boxed{1}\end{bmatrix} \\ \texttt{REST} & \mathit{phrase}\begin{bmatrix}\texttt{HEAD} & \boxed{1}\end{bmatrix}\end{bmatrix}$ (b) $\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{FIRST} & \mathit{non\text{-}head}\begin{bmatrix}\ \end{bmatrix} \\ \texttt{REST} & \mathit{phrase}\begin{bmatrix}\texttt{HEAD} & \boxed{1}\end{bmatrix}\end{bmatrix}$

[14] This is actually not quite correct. List representations in the form of attribute value pairs normally consist of a feature structure of the type *list* (or *cons*) with two attributes `FIRST` and `REST`, where the value of the latter can be of the type *list* or *nil*. Oliva, however assumes that the node itself may contain additional features, specifically `SYNSEM`. From this point of view, it is actually questionable whether his recoding is not just a relabelling of traditional daughters attributes, with the (superfluous) attribute `DTRS` eliminated of course.

[15] We ignore irrelevant attributes such as `SYNSEM` etc.

However, in these specifications the notion of a head does not have a straightforward definition anymore. Is it the phrase which is typed as *head*, in which case the head notion corresponds to the traditional label of the HEAD-DTR, but is not defined in the (b) variant of the HFP. Or is it equivalent to the structure whose HEAD value is identified with the mother (analogously to the traditional HFP), in which case the uniqueness of the head is not given anymore in the (a) part.

Similar to the HFP, also the SCP is split up into four different, disjoint cases, which are again aligned with the cases of the HFP. Those two of the possible resulting configurations given in (325) we can practically ignore, since they are basically identical to head-final head-complement- and head-adjunct-schemata with renamed daughters labels.

(325)
$$\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \\ \texttt{FIRST} & {}_{comp}\boxed{3}\,[\;] \\ \texttt{REST} & {}_{phrase}\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \cup \left\{ \boxed{3} \right\} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \\ \texttt{FIRST} & {}_{adj}[\;] \\ \texttt{REST} & {}_{phrase}\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

The two components which are actually 'doing the job' are the configuration in which the head-initial verb is combined with the 'rest' (326) and the clause- or list-final element, which can be seen as an empty terminal or as an end-of-list symbol (327).

In the phrasal configuration (326) two pieces of information are encoded, which make a lexical rule of the [Kiss and Wesche 1991] style unnecessary. Firstly, one clause of Oliva's SCP is realized, which specified that in head-initial configurations the SUBCAT value of the head is identified with a HEAD feature HEAD-SUBCAT, underlying the HFP. Secondly, the HEAD features of the head are identified with those of the mother and of the REST attribute.

(326)
$$\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{FIRST} & {}_{head}\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SUBCAT} & \boxed{2} \end{bmatrix} \\ \texttt{REST} & {}_{phrase}\begin{bmatrix} \texttt{HEAD}\ \boxed{1} \begin{bmatrix} \texttt{HEAD-SUBCAT}\ \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The second element required in this analysis is the clause-final empty terminal or an end-of-list symbol encoded as a typed feature structure. It takes the HEAD-SUBCAT value off the carrier feature and makes it its own SUBCAT feature, same as the empty terminal in the [Kiss and Wesche 1991] above adopted the LOCAL value from the DSL feature. This SUBCAT feature can then be reduced by the other clauses of Oliva's SCP, from right to the left.

$$(327)\qquad {}_{nil}\begin{bmatrix} \texttt{HEAD} & \begin{bmatrix}\texttt{HEAD-SUBCAT}\ \boxed{1}\end{bmatrix} \\ \texttt{SUBCAT} & \boxed{1} \end{bmatrix}$$

If we now take a closer look at the head-initial configuration in (325), there are two aspects which are somewhat at odds with the traditional spirit of the HFP and SCP, making this structure rather exceptional. As just mentioned, it is quite unusual that the head features of a mother node are identified with both the head and the non-head daughter. It is even more unusual, however, that the SCP neither reduces the subcategorization of the head nor identifies it with that of the mother, but rather leaves the mother's value completely unspecified, while it is not instantiated by a `SUBCAT` value coming from the `REST` daughter.

To some extend this is apparently intended, since this underspecification *a priori* allows the head to appear in practically any position. Since there are languages in which this may well be the case, such as Czech, this is in principle a matter of parametrizing the `SUBCAT` value of the mother to determine the actual position of the verbal head. However, oddly enough, if we associate the notion of head with the one daughter providing the `SUBCAT` value and the `HEAD` value of the mother, the `REST` phrase comes considerably closer to this definition than the `FIRST` value typed as *head*.

If we then compare the approach of [Oliva 1992] with the analysis suggested by [Kiss and Wesche 1991], the crucial differences (modulo other side effects) appear to be found in the specification of the grammatical relations between the clause-initial verb and its sister constituent, and in the devices employed to convey the `SUBCAT` information to the clause-final empty terminal. [Kiss and Wesche 1991] attempt to maintain the Subcategorization Principle and assign a head-complement relation to the clause-initial finite verb and its right-hand sister, but they can do so only by exploiting the full power of lexical rules to re-encode the subcategorization of the finite verb. [Oliva 1992] more or less ignores the standard Subcategorization Principle locally for the head-initial configuration altogether and simply does not make any commitment to the relation between the finite verb and rest of the clause. In the case of [Kiss and Wesche 1991], the carrier of the `SUBCAT` information is a non-local feature which has to be artifically constrained to a local domain. In the case of [Oliva 1992], it is the `HEAD` feature itself transporting the information, again at the price of locally suspending the effects of the standard HFP, which has the effect that in this approach the notion of a head becomes considerably less perspicuous.

Frank 1994 presents a rather sophisticated analysis in which she tries to combine the trace mechanism approach of [Kiss and Wesche 1991] with our own approach to functional heads described in the previous chapter, thus describing an analysis which models the G&B-oriented approach of [Haider 1993] within HPSG. Apparently, without being aware of the fact she also incorporates into this analysis some aspects of [Oliva 1992], specifically the employment of the HFP and some `HEAD` feature as a carrier for the subcategorization information. Although this analysis chronologically should be sorted in after our own analysis of V-initial structures, as most of the fundamental aspects of our approach had already been published in [Netter 1992], it is conceptually much closer to the other analysis by [Kiss and Wesche 1991] and [Oliva 1992] presented in this subsection. As will become clear very shortly, the approach of [Frank 1994] (like our own outlined below) is an analysis of V-initial which builds on our notion of functional heads, but which branches off into quite a different direction by attempting to model the effects of head movement in HPSG.

The objective of Frank is to develop an analysis for V-initial which implements the notion of head-movement in HPSG, but which improves on some of the deficiencies of the approach by [Kiss and Wesche 1991] in the same way as [Haider 1993] improves on traditional accounts of head-movement within the G&B framework. Her main criticism of these approaches is that the relation between the moved head and the clause final trace typically does not involve the sharing of crucial information as it is normally assumed for antecedent–trace relations.

Typical G&B accounts attempt to model head or V-initial movement by structures such as (328), in which two quite different categories, I^0 and V^0 are co-indexed.

(328) $[_{IP}\ \mathrm{SpecI}\ [_{\overline{\mathrm{I}}}\ \mathrm{I}^0_i\ [_{VP}\ ...\mathrm{V}^0_{e_i}\ ]]]$

In a similar way, she argues, the approach by [Kiss and Wesche 1991] fails to capture the relationship between the antecedent verb and the clause-final terminal, since there is practically no relationship between the `LOCAL` value of the antecedent (after it has been derived by a lexical rule) and the one of the trace.

In the same way as Haider suggests an alternative to (328) by postulating a functional projection FP with a variable category F^0 which can be filled by a complementizer or the finite verb (329), Frank suggests to treat the finite verb in a clause-initial position as a functional head in our sense.

(329) $[_{FP_{(Vfin)}}\ \mathrm{SpecF}\ [_{\overline{\mathrm{F}}_{(Vfin)}}\ \mathrm{F}^0_{(Vfin)_i}\ [_{VP}\ ...\ \mathrm{V}^0_{e_i}\ ]]]$

Although Frank shows that the [Kiss and Wesche 1991] approach involving a lexical rule could be considerably improved by incorporating our notion of functional completeness, she proceeds along some different lines and dispenses with the lexical rule altogether in favour of an analysis based on underspecification. The central idea of this analysis is that the finite verbs are heads in both the initial and final position, but have a quite different structure which is derived from a common underspecified representation.[16] All verbs crucially have an abstract representation in common, in which their subcategorization frame is represented as a HEAD feature. In normal substantive lexical categories this frame is mapped onto the revised valence features SUBJ and COMPS, in functional categories no such mapping or coindexation takes place.

V-initial verbs (330) are then functional heads which take as a complement the *Mittelfeld* headed by an empty terminal. The entire HEAD value of the category is co-indexed with the HEAD value of the complement, as well as with a TO-BIND value of DSL. The co-indexation of the entire HEAD value marks finite verbs as 'secondary' functional categories in contrast to primary functional categories, such as complementizers, in which only the MAJOR feature is shared. Note that this also implies that secondary functional categories take complements which have to be marked as functionally complete. The TO-BIND value of DSL serves to cancel a corresponding INHERIT value eventually arising from the sister complement, thus satisfying the standard Non-Local Principle.

(330)

$$ {}_{\textit{sec-func-cat}}\left[\begin{array}{ll} \texttt{LOC|CAT} & \left[\texttt{HEAD}\ \boxed{1}\left[\begin{array}{l}\texttt{MIN|FC}\ +\\ \texttt{SUBCAT}\ \langle ...\rangle\end{array}\right]\right]\\ \texttt{COMPS} & \left\langle\left[\texttt{LOC|CAT|HEAD}\ \boxed{1}\right]\right\rangle \\ \texttt{NON-LOC} & \left[\texttt{TO-BIND|DSL}\ \boxed{1}\right]\end{array}\right] $$

The clause-final empty node (331) is then analyzed as a substantive category, which is different from other substantive categories in that it is being marked as functionally complete and as having a non-empty DSL value co-indexed with its HEAD feature. As with all other substantive categories its valence features are co-indexed with the head feature SUBCAT.

(331)

$$ {}_{\textit{empty-lex-cat}}\left[\begin{array}{ll} \texttt{LOC|CAT} & \left[\begin{array}{l}\texttt{HEAD}\ \boxed{1}\left[\begin{array}{l}\texttt{MIN|FC}\ +\\ \texttt{SUBCAT}\ \textit{cons}(\boxed{2},\boxed{3})\end{array}\right]\\ \texttt{SUBJ}\ \ \boxed{2} \\ \texttt{COMPS}\ \ \boxed{3}\end{array}\right]\\ \texttt{NON-LOC} & \left[\texttt{INHERIT|DSL}\ \boxed{1}\right]\end{array}\right] $$

As has been already mentioned above, the crucial point about these structures is that it is *not* the DSL feature which is the primary carrier of the subcategorization information, but rather the HEAD feature. At least, the function of the DSL feature

[16]We ignore in the following those aspects which are basically common with our own analysis and which will be also discussed below in section 5.3.

is not primarily to carry this information, but rather to establish an antecedent-trace relation between the finite verb and the empty terminal. Since the head feature already conveys the same information, it appears that the DSL would be almost redundant, if it did not serve to distinguish empty substantive categories from non-empty ones, which must have an empty DSL value and which are marked as functionally incomplete. Thus, in some respect only this latter function, which helps to block the combination of two identical (non-empty) finite verbs in an initial and a final position, still motivates the presence and application of a trace mechanism in Frank's approach. For other, more practical purpose, it seems that the transportation mechanism of the HFP would have been sufficient.

Note finally that in spite of Frank's claim to dispense with lexical rules the different representation of V-initial and V-final verbs would still make it necessary to derive two entries at compile time if type inference could not be employed at runtime.

5.2.5 Summary

In this section we compared different methods and approaches to capture the relationship between finite verbs in clause-initial and final position as well as the interaction with the structure of the *Mittelfeld* in German clauses. We showed that each of these methods alternatively had to assume either (a) a modification of the standard adjunction or the subcategorization mechanism, (b) a non-uniform structural representation for V-initial and V-final clauses, or (c) some additional devices which accounted for the percolation of the V-initial subcategorization frame to some clause-final empty terminal.

The case of (a) was illustrated by the approach of [Pollard 1991], who postulates a flat clause structure. The serious problems arising from this approach are addressed by [Kasper 1994] who suggests to invoke universal constraints to cope with recursive adjunction and by [Nerbonne 1994b], who proposes a lexical treatment for non-local dependencies to eliminate spurious ambiguities in the case of partial VP fronting. Approaches falling under class (b) so far have not been suggested in the HPSG literature, but can be found in some related form in the categorial grammar framework. The crucial advantage of both of these two classes of analyses is that they can do without empty terminals.

In contrast to these approaches a number of suggestions have been developed to analyze German clause structure by a binary right-branching recursion. All of these approaches have in common that they postulate a clause-final empty terminal, which is connected with the clause-initial finite verb by some special mechanism. The carrier features suggested to transport the subcategorization frame to this empty node are either a special DSL feature, in the case of [Kiss and Wesche 1991], some HEAD feature, in the case of [Oliva 1992] or both, in the case of [Frank 1994].

In our own analysis below, which is probably most closely related to the one by [Frank 1994] as it also builds on our concept of functional heads, we will show that it is possible to do without any of these extra mechanisms, that we can maintain the standard principles, that we do not need two different lexical entries for finite verbs and that we can even maintain a substantial part of our own principles governing the distribution of functional categories in the area of nominal categories.

5.3 Verbal Functional Heads

In the following section, we present an alternative approach to the above analyses, which closely builds on the notion of functional heads as we developed in section 4.3 above. We show how the assumptions developed for nominal phrases carry over to the verbal domain and how they can be extended to account for the variations of the finite verb position in German.

Before we go on, we briefly summarize some of the basic assumptions in which we differ from the above-mentioned approaches as well as the most important generalizations that we attempt to account for.

- We assume a (uniform) right branching binary structure for the *Mittelfeld*, among other reasons because such a structure greatly simplifies the description of mixed adjunct and complement sequences. The assumption of such a structure could also contribute to capture phenomena such as scope, binding or coordination.
- We reject the assumption of two different types of lexical entries for finite verbs in a V-final and a V-initial position.
- Since the relation between an initial finite verb and a clause final empty head is clearly a local relation, which should be restricted to head domains, we refuse the assumption of mechanisms designed for the analysis of non-local dependencies.
- There is an obvious complementary distribution between complementizers, such as *daß*, *ob*, *wenn*, etc., and the clause-initial position of the finite verb. Any theory which fails to account for this mutual dependency in a non-arbitrary way does not even come close to an explanation of the phenomenon.
- We assume that in an analysis of finite verb position the basic principles and rules of HPSG should be maintained even without the introduction of extra mechanisms.

In order to account for these generalizations we will adopt one assumption from the 'movement' analysis for V-initial, namely that there is a clause-final empty node which functions as the head of the clausal structure. This assumption arises from the generalization about branching direction and head-position which cannot be circumvented without giving up the HPSG's subcategorization mechanism. To establish the relation between this empty node and the finite verb we will make use of the existing *subcategorization* mechanism of HPSG rather than introducing new principles or modifying the existing ones.

In terms of the three parameters, referred to in the beginning of section 5.2, we will thus maintain the assumption of a uniform structure as well as the standard mechanisms and principles, but we will abandon the assumption that the finite verb in a clause-initial position is to be analyzed as a head *in situ*.

We begin the presentation of our approach with some basic assumptions about the lexical structure of substantive verbal categories and the representation of functional categories such as complementizers, analogous to the one given for nominal categories. We then proceed to show how the notion of functional completeness can be viewed as a property which can be lexically or structurally anchored, and how this notion can be employed in the analysis of V-initial clauses. For this purpose we introduce an empty functional head which can take the finite verb as a complement. To simplify the computational load, as well as to extend the analysis to constructions with a more relaxed word order, we also give an alternative representation for the empty functional head, which is also more compatible with the spirit of the revised HPSG Valence Principle. We finally show how separable verb prefixes can be easily accommodated with the analysis.

5.3.1 V-Final Clauses with Complementizers

The distinction that we set up for the description of nominal phrases more or less directly carry over to the specification of substantive and functional categories in the verbal domain, at least as long as we are concerned with V-final clauses with complementizers.

The description of phrasal (i.e. clausal) and lexical categories is based on the type definitions specifying the `MAJOR` features of verbal categories. These definitions were already introduced in section 4.3.1 and are briefly repeated here for ease of reference.

(332)

$$\begin{array}{lcl} \textit{verb} & \equiv & \textit{cat} \;\wedge\; \left[\texttt{LOC}|\texttt{CAT}|\texttt{HEAD}|\texttt{MAJOR}\left[\texttt{POS}\begin{bmatrix}\texttt{N} & - \\ \texttt{V} & +\end{bmatrix}\right]\right] \\ \text{part}(\textit{verb}) & \equiv & \textit{fin-verb},\ \textit{non-fin-verb} \end{array}$$

(333) $$\textit{fin-verb} \;\equiv\; \textit{verb} \;\wedge\; \left[\texttt{LOC}|\texttt{CAT}|\texttt{HEAD}|\texttt{MAJOR}\left[\texttt{AGR}\begin{bmatrix}\texttt{NUM} & [\;] \\ \texttt{PERS} & [\;]\end{bmatrix}\right]\right]$$

(334) $\textit{tensed-verb} \equiv \textit{fin-verb} \wedge \left[\texttt{LOC}|\texttt{CAT}|\texttt{HEAD}|\texttt{MAJOR} \begin{bmatrix} \texttt{TENSE} & \textit{tense}[\ ] \\ \texttt{MOOD} & \textit{mood}[\ ] \end{bmatrix} \right]$

(335) $\textit{imper-verb} \equiv \textit{fin-verb} \wedge \left[\texttt{LOC}|\texttt{CAT}|\texttt{HEAD}|\texttt{MAJOR} \begin{bmatrix} \texttt{TENSE} & *\textit{undef}* \\ \texttt{MOOD} & *\textit{undef}* \end{bmatrix} \right]$

While these type definitions cut across verbal categories irrespective of their lexical or phrasal status or their level of projection, we also have to define the lexical types of substantive verbal categories. In contrast to the nominal substantive categories, for which we assumed that they can be underspecified with respect to functional completeness, we define all lexical entries of finite verbs to be intrinsically specified as functionally incomplete.

This specification is motivated by the assumption that whenever a finite verb in German is involved in the formation of a maximal phrasal projection, be it a main or a subordinate clause, it is bound either to the presence of some *lexical* functional head or to some *structural* marking of the finite verb in the sense that the verb does not occur in its "normal" clause-final position. The former configuration we find in combination with complementizers introducing subordinate or generally V-final clauses. The latter configuration, i.e., V-initial constructions, we will address below.[17]

We will accordingly augment our type lattice by the following definitions for the substantive lexical entries of tensed finite verbs and imperatives.[18]

(336) *tensed-V-lex* $\equiv$ *tensed-verb* $\wedge$ *f-incompl-cat*

(337) *imper-V-lex* $\equiv$ *imper-verb* $\wedge$ *f-incompl-cat*

The analogue of these categories in the nominal domain is the functional incomplete lexical nominal category, referred to as *sg-count*. Similarly, the other definitions required for the description of tensed V-final clauses practically fall out

[17]This generalization is crucially constrained to the case for finite constructions. Note that there are plenty of constructions like (i)/(ii), where non-finite verbs project into full maximal phrase without any specifiers or structural re-organisation.

(i) Den Rasen nicht betreten

(ii) Saarländischer Finanzminister zurückgetreten

Whether non-finite verb forms should be assumed to be altogether unmarked in the lexicon we want to leave open here. The answer to this question will presumably very much depend on the type of analysis given to constructions involving "non-finite" complementizers such as *um* or *anstatt*, which could be treated as functional heads or alternatively as prepositions taking a non-finite clausal complement.

[18]We will however henceforth mostly ignore the branches specifying non-finite and imperative categories and projections, since they either do not play a role or are subsumed by other relevant types.

from the cross-classification of the verbal category types with the category neutral definitions for maximal projections, functional heads etc.

Complementizers, such as *daß*, *ob*, *wenn*, etc., are defined by intersecting the specific type for functional heads *func-cat* with the type definition for tensed verbs. Since complementizers do not occur together with imperative forms of verbs they cannot simply inherit from finite verbs, but have to be combined with the more specific *tensed-verb* type.

(338) $\textit{fin-cpl} \;\equiv\; \textit{tensed-verb} \;\wedge\; \textit{func-cat}$

Remember that the definition of *func-cat* requires the complement of the functional head to be both functionally incomplete and fully saturated. Accordingly, the verbal complement selected by a complementizer, i.e., the body of a finite V-final clause is defined by the intersection of tensed verbs and functionally incomplete saturated projections (339). The maximal projection of tensed finite clauses are functionally complete and fully saturated categories *max-cat* intersected with the tensed verbal category (340).[19]

(339) $\textit{tensed-s} \;\equiv\; \textit{tensed-verb} \;\wedge\; \textit{f-inc-sat-cat}$

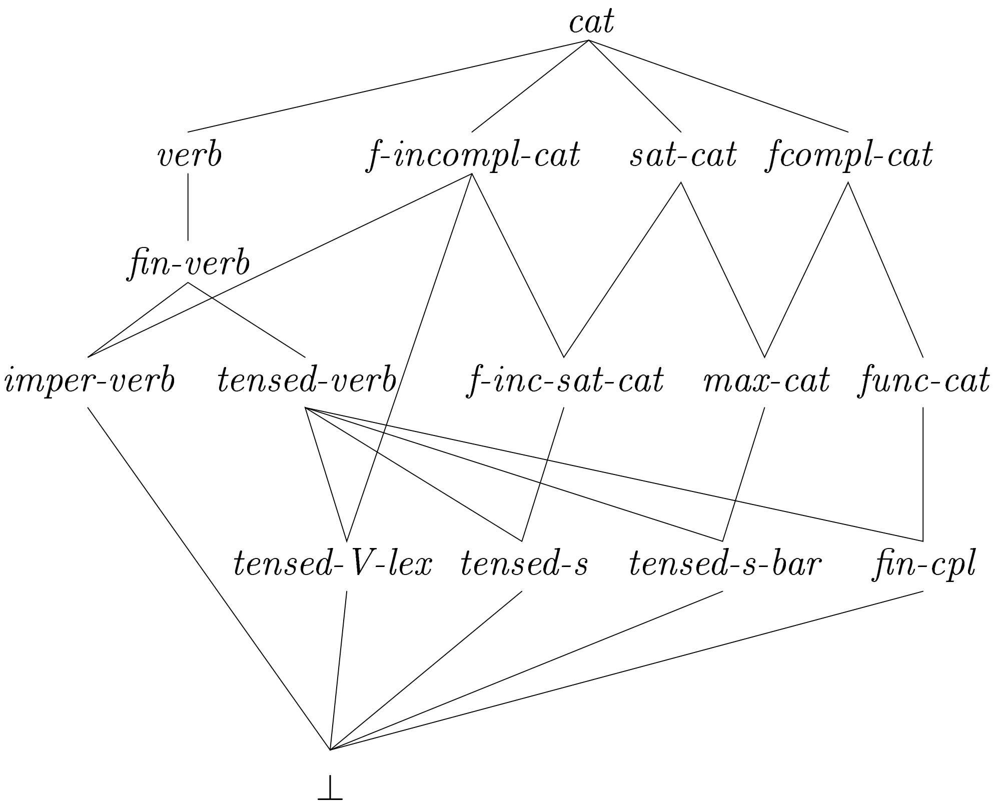
(340) $\textit{tensed-s-bar} \;\equiv\; \textit{tensed-verb} \;\wedge\; \textit{max-cat}.$

The expanded entry for a complementizer with a `SUBCAT` value resolved to *tensed-s* is given in (341). This entry also contains the feature `SPEC` under its minor head features which we already employed in the description of determiners and whose more specific function will be explained below.

(341)
$$
{}_{\textit{fin-cpl}}\left[\begin{array}{ll}
\text{LOC|CAT} & \left[\begin{array}{ll}
\text{HEAD} & \left[\begin{array}{ll}
\text{MAJOR}\ \boxed{1} & \left[\begin{array}{ll}
\text{POS} & \left[\begin{array}{l} \text{N}\ + \\ \text{V}\ - \end{array}\right] \\
\text{AGR} & \left[\begin{array}{l} \text{NUM} \\ \text{PERS} \end{array}\right] \\
\text{TENSE} & \textit{tense}\left[\ \right] \\
\text{MOOD} & \textit{mood}\left[\ \right]
\end{array}\right] \\
\text{MINOR} & \left[\begin{array}{ll} \text{FCOMPL} & + \\ \text{SPEC} & + \end{array}\right]
\end{array}\right] \\
\text{SUBCAT} & \left\langle {}_{\textit{tensed-s}}\left[\begin{array}{ll}
\text{LOC|CAT} & \left[\begin{array}{ll}
\text{HEAD} & \left[\begin{array}{ll} \text{MAJOR} & \boxed{1} \\ \text{MINOR} & \left[\text{FCOMPL}\ -\right] \end{array}\right] \\
\text{SUBCAT} & \langle\ \rangle
\end{array}\right]
\end{array}\right] \right\rangle
\end{array}\right]
\end{array}\right]
$$

[19] Again, we choose the names *S* and *S-bar* mainly for lack of an accepted alternative terminology, rather than because we want to commit ourselves to some version of $\overline{\text{X}}$-theory.

(342)



Note that apart from the `MAJOR` features the definition of complementizers is exactly analogous to the definition of determiners. This is hardly surprising as the last three type definitions are formed in a full analogy to their corresponding types in the nominal domain, viz., *det*, *np* and *dp*, respectively. This analogy to the nominal constructions becomes even clearer if we compare the respective sections of the type lattices with each other. If we compare the type lattice in (342) with its counterpart in the nominal domain given in section 4.4, we find practically the same graph geometry with the differences being constrained to the labels of types inheriting from *verb* and *noun* respectively.

The analysis of V-final finite constructions with a complementizer then basically follows the same derivation pattern that is applied in the case of nominal constructions with a specifier or determiner. Again, the only rule schema required is a binary branching *head-complement-schema*. This schema will first combine the clause final verb with its complements to its left, until this functionally incomplete projection is fully saturated, and can be then applied to establish the head-complement relation between the complementizer as the functional head of the clause and the saturated verbal projection.

One aspect of this analysis needs further clarification. So far, we had assumed a generalization for German word order to the effect that two linear precedence constraints will restrict heads which are marked [V $-$] to occur in a head-initial position, while those which are marked [V $+$] are assigned a head-final position. Given the definitions and classifications introduced above, the latter constraint will not yield the correct results and has to be modified. Since substantive and functional verbal heads appear in a head-final and head-initial position respectively, we would get conflicting results if we based the LP constraints exclusively on the major part of speech classification. On the other hand, we also cannot base the LP generalization on the property of functional completeness, since we will assume below that V-initial clauses contain a functionally complete verbal head in a head-final position.

However, while functional completeness was never really limited to functional categories, there is another minor property which was assigned only to a subset of functional categories, viz., the feature [SPEC $+$]. Thus, the most adequate revision of the LP constraints for our purposes appears to be a generalization that categories which are marked as specifiers, i.e., [SPEC $+$], have to occur in a head-initial position in German. This constraint takes priority over the POS-specific LP constraints, which continue to hold, albeit conditional on the SPEC values of a head. In the case of nominal phrases the two different constraints will always yield the same linear order. In the case of verbal projections we have to multiply out the constraints, so that heads marked as [SPEC $-$], [V $+$] follow the arguments or adjuncts, whereas those with [SPEC $+$], [V $+$] have to precede them.

In some respect, this reformulation somewhat weakens the thrust of the argument that functional heads must have the same MAJOR features as their complements, because they occur in a comparable position as the corresponding substantive heads. On the other hand, the LP constraints as they are given now group together functional heads of a certain type, viz., those that are lexical overt and that have to occur in a peripheral position. Above all, it has to be kept in mind that the SPEC feature has not been motivated exclusively or even primarily on the basis of word order regularities. The value of this attribute plays an equally important role for the problem of adjunction, allowing us to constrain potential adjunction sites across the board on the basis of the SPEC feature, rather than on the basis of functional completeness. Both determined nominal phrases and V-final clauses with complementizers do not allow adjunction to the left of their specifiers, while it is not at all excluded that adjuncts will be combined with functionally complete categories, as we showed in section 4.5 above and as will become clear below.

Note by the way, that the approach of [Frank 1994] also assumes that LP constraints should not be based on part of speech distinctions, but rather on minor properties cutting across different major categories. One of the nicest features of her analysis is a straightforward generalization that functionally complete heads occur in a left-peripheral position, whereas functionally incomplete categories are head-final. Nevertheless, this apparent advantage of her analysis has be bought at the price of postulating two different lexical entries for finite verbs, which are either functionally complete in V-initial clauses or incomplete in V-final constructions. Also, it is not quite clear how the relationship between a topicalized phrase and the finite verb in a V-second construction has to be seen, since relative to the filler the functionally complete verb will be head-final.

Still, leaving aside the issue of linear precedence, there are still some generalizations that simply fall out from our treatment of complementizers as functional heads. For example, in our account complementizers have to appear in a (left-) peripheral position, which follows from the fact that they only combine with saturated verbal projections and adjuncts cannot adjoin to specified constructions. Secondly, projections of finite verbs with the verb in clause-final position but without complementizers do not occur on their own, since they are functionally incomplete and all maximal projections are required to be functionally complete. Another consequence of our approach, viz., the complementary distribution of V-initial position and complementizers, will be explained in the next section.

5.3.2 V-Initial Clauses and Non-Head Non-Movement

While in the case of complementizers we have an overt lexical item which can function as the functional head of the clause and which can carry the feature marking the clausal projection as functionally complete, there is no such obvious lexical head in the case of V-initial clauses. On the other hand, we observed that in the absence of such an overt functional head the finite verb has to appear in a clause-initial position. As a natural conclusion, we could assume that the intrinsic functional incompleteness of the finite verb in this case is compensated on a *structural* rather than a lexical level.

One approach to describe such a structurally induced functional completeness is to modify the `DSL` approach by [Kiss and Wesche 1991] and to assume that the lexical rule which derives the clause-initial finite verb also marks this head as functionally complete. From such an extension, we would gain the generalization that V-initial and complementizers are in a complementary distribution. This direction is basically pursued by [Frank 1994], as described above. However, we take the position that one would clearly do better if the lexical rule or duplicate lexical entries can be avoided altogether, even if an empty clause-final terminal node still has to be assumed.

We therefore propose to treat this empty terminal itself as the functional head of the clause, marking it as functionally complete and sharing its major features with the functionally incomplete finite verb. The lexical structure of this empty functional head will share many of the properties of lexically overt functional categories, viz., determiners and above all complementizers, but it also will have to differ in some crucial aspects.

We will assume that this empty node has in common with complementizers that it subcategorizes for a functionally incomplete verbal category, however in this case the complement will be constrained to a non-phrasal lexical category. Exactly like complementizers the empty head will adopt the `MAJOR` features from the verbal category, while introducing itself the `MINOR` head feature which marks the clause as f-complete. Under this assumption, the clause-initial finite verb will not function as the `HEAD-DTR` of the clause anymore, but simply as a complement daughter and direct argument of the functional head. However, all of its relevant `MAJOR` properties (e.g., tense, mood, agreement features etc.) will still be passed on through the empty head to the maximal projection of the clause. In contrast to complementizers this empty head has to be marked as `SPEC`-, which will admit the adjunction of adverbial phrases at any level of projections from the empty head, and which will also satisfy the relevant LP constraints requiring verbal heads with this marking to be head-final.

What we have fixed by these assumptions so far is only the grammatical relation between the finite verb and the empty functional head. Still missing is an account for the transmission of the finite verb's `SUBCAT` frame to the empty head, as well as the means to determine the position of the finite verb. As we have seen above in section 5.2, the first problem is typically solved by means of a carrier feature transporting the finite verb's subcategorization information. Since we assume a head-complement relation between the finite verb and the functional head, the carrier feature in our case is quite naturally the feature `SUBCAT`. Thus, if the verb functions as a complement, the empty functional head must subcategorize on the one hand for the finite verb itself, but also for all the complements that this finite verb subcategorizes for.

In [Netter 1992] we therefore suggested to capture the relation between the empty head and finite verb through the standard subcategorization mechanism and to determine the empty head's `SUBCAT` frame by means of a list operation resulting in a concatenation of the subcategorization for the finite verb and its own complements.

In (343) the `SUBCAT` value of the functional head is specified as the result of a functional constraint **snoc**. This function, which can be seen as the "inverse" of the standard **cons** operation or as a specific case of **append**, has to be defined such that it appends a single element at the end of a list, yielding as a result a list. In this case, the function will produce as its result a list which is obtained

(343) Empty Verbal Functional Head

$$
{}_{eps}\left[\begin{array}{ll}
\mathtt{LOC} & \left[\begin{array}{ll}
\mathtt{CAT} & \left[\begin{array}{ll}
\mathtt{HEAD} & \left[\begin{array}{ll}
\mathtt{MAJ}\ \boxed{1} & \left[\begin{array}{ll}
\mathtt{POS} & \left[\begin{array}{l}\mathtt{N}\ - \\ \mathtt{V}\ + \end{array}\right] \\
\mathtt{AGR} & [\ ] \\
\ldots &
\end{array}\right] \\
\mathtt{MIN} & \left[\begin{array}{ll}\mathtt{FCOMPL} & + \\ \mathtt{SPEC} & - \end{array}\right]
\end{array}\right] \\
\mathtt{SUBCAT}\ \mathbf{snoc} & \left({}_{vfin}\left[\begin{array}{ll}
\mathtt{LOC} & \left[\begin{array}{ll}
\mathtt{CAT} & \left[\begin{array}{ll}
\mathtt{HEAD} & \left[\begin{array}{ll}\mathtt{MAJ} & \boxed{1} \\ \mathtt{MIN} & \left[\mathtt{FCOMPL}\ -\right]\end{array}\right] \\
\mathtt{SUBCAT} & \boxed{3}
\end{array}\right] \\
\mathtt{LEX} & + \\
\mathtt{CONT} & \boxed{2}
\end{array}\right]
\end{array}\right], \boxed{3} \right)
\end{array}\right] \\
\mathtt{CONT} & \boxed{2}
\end{array}\right]
\end{array}\right]
$$

from appending the feature structure with the finite verb's SYNSEM value to the end of the list which represents the SUBCAT value of this verb.

We illustrate the function by the example of a transitive finite verb which subcategorizes for two nominal complements. If we instantiate the first argument of **snoc** with the lexical entry of such a verb, the functional constraint specifying the SUBCAT value of the empty node takes the form given in (344).

$$
(344)\quad \mathbf{snoc}\left({}_{vfin}\left[\begin{array}{ll}
\mathtt{LOC|CAT} & \left[\begin{array}{ll}
\mathtt{HEAD} & \left[\begin{array}{ll}\mathtt{MAJ} & \left[\begin{array}{l}\mathtt{N}\ - \\ \mathtt{V}\ + \end{array}\right] \\ \mathtt{MIN} & \left[\mathtt{FCOMPL}\ -\right]\end{array}\right] \\
\mathtt{SUBCAT} & \boxed{3}\left\langle {}_{np\text{-}acc}[\ ], {}_{np\text{-}nom}[\ ]\right\rangle
\end{array}\right]
\end{array}\right], \boxed{3} \right)
$$

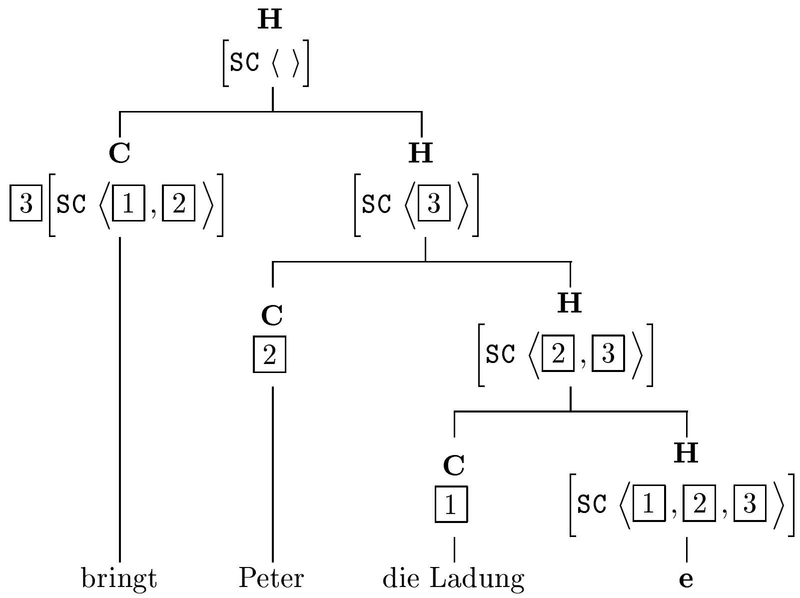
As the result of the evaluation of this function we obtain a list of three elements consisting of the two arguments of the verb and the verb itself:

$$
(345)\quad \left\langle \boxed{1}{}_{np\text{-}acc}[\ ],\ \boxed{2}{}_{np\text{-}nom}[\ ],\ {}_{vfin}\left[\begin{array}{ll}
\mathtt{LOC|CAT} & \left[\begin{array}{ll}
\mathtt{HEAD} & \left[\begin{array}{ll}\mathtt{MAJ} & \left[\begin{array}{l}\mathtt{N}\ - \\ \mathtt{V}\ + \end{array}\right] \\ \mathtt{MIN} & \left[\mathtt{FCOMPL}\ -\right]\end{array}\right] \\
\mathtt{SUBCAT} & \left\langle \boxed{1}, \boxed{2} \right\rangle
\end{array}\right]
\end{array}\right] \right\rangle
$$

The operation underlying this analysis is technically very similar to the verb-raising or rather SUBCAT-raising operation as it was suggested for the description of clause union phenomena with verb clusters involving auxiliaries and modal verbs by [Hinrichs and Nakazawa 1989]. The lexical entry of the empty functional head differs from these clause union verbs insofar as the SUBCAT-raising operation cannot be applied recursively. This ensures that the relation between the finite verb and the empty head will always be constrained to a local head-domain, a property which did not come out naturally in the approach of [Kiss and Wesche 1991]. Secondly, while the governing verb in a verb cluster may also function as the semantic head of a clause, this will never be the case for an empty functional head, which always fully adopts the CONTENT value of the finite clause through co-indexation.

Most importantly, while auxiliaries and modals *first* combine with their verbal complement and only then with the other, "raised" arguments, the empty functional head must combine the finite verb as the very last argument, while it consumes the other complements first. To implement this regularity, one can make the simplified assumption that a binary head-complement schema will reduce the SUBCAT list from left to right by one element at a time, which would mean that we do not need any other devices—rule instances, principles or whatever—to derive V-initial clauses. Above all, we will not need any other linear precedence constraints to determine the branching direction of the clausal structure beyond the ones postulated for the V-final clauses with complementizers. The relevant LP constraint applying in this case will require that a HEAD-DTR specified as [V +] and [SPEC −] has to follow a COMP-DTR or an ADJ-DTR. Since the finite verb is the last complement on the list it will automatically appear at the left periphery of the structure in a clause initial position, as illustrated in (346).

(346)



5.3.3 V-initial and Free Word Order

The way it has been just specified above, the analysis of V-initial will very much rely on the assumption that subcategorization frames are processed strictly from left to right and that the finite verb occurs as the very last element on the subcategorization list. However, such an assumption cannot be easily maintained anymore, if we want to account for a more relaxed word order in the domain of nominal and prepositional complements. In a binary branching structure such a relaxed word order will most likely entail that the elements on the `SUBCAT` can be taken off in any order. To ensure under these conditions that the finite verb still appears to the very left of the complements and adjuncts, it will be necessary to introduce some additional constraints.

One possibility would be to resort to non-local LP rules, as they were suggested for example by [Engelkamp et al. 1992] or [Oliva 1992], whose scope goes beyond the strictly local domain of trees of depth one. In a binary branching approach such LP rules could be employed for example for determining the "soft" linear precedence constraints governing the order of complements and adjuncts among themselves. It is doubtful, however, whether they should also provide the device for determining the position of the finite verb, since the relative order of nominal and prepositional complements in the *Mittelfeld* appears to be of a quite different nature than the "hard" constraints determining position of the finite verb. Thus, even if we suggested to exploit the subcategorization mechanism to obtain the correct dependencies, we are still somewhat reluctant to treat the finite verb as a complement fully on a par with others.

As an alternative to determining the position of the finite verb (relative to other complements) either by the order of complements on the `SUBCAT` list or by means of non-local LP constraints, we suggest a more refined classification of complements along the lines of the revised Valence Principle. This means that we separate out the selection of the finite verb and treat it through another valence feature, say `VCOMP`, which occurs on a par with the already existing valence features `SUBJ` and `COMPS`.[20] The revised entry for the empty functional head will then take the form in (347).

The introduction of this additional valence feature will of course have to be matched by an appropriate extension of the Valence Principle and it will require a further subtype or instance of the head-complement schema, which will cover the combination of an (initial) finite verb with a head-final head. Through this schema the actual position of the finite verb can be determined by setting the values of the other valence features to be empty-listed (348), so that all other

[20]It should be kept in mind that the introduction of these two features is also partly motivated by the fact that the treatment of configurational or "hard" word order regularities will be simplified by this distinction.

(347) Empty Verbal Functional Head (revised version)

$$
{}_{eps}\left[\begin{array}{ll}
\mathtt{LOC|CAT} & \left[\begin{array}{ll}
\mathtt{HEAD} & \left[\begin{array}{ll}
\mathtt{MAJ} & \boxed{1} \\
\mathtt{MIN} & \left[\begin{array}{ll} \mathtt{FCOMPL} & + \\ \mathtt{SPEC} & - \end{array}\right]
\end{array}\right] \\
\mathtt{VAL} & \left[\begin{array}{ll}
\mathtt{VCOMP} & \left\langle {}_{vfin}\left[\begin{array}{ll}
\mathtt{LOC|CAT} & \left[\begin{array}{ll}
\mathtt{HEAD} & \left[\begin{array}{ll} \mathtt{MAJ} & \boxed{1} \\ \mathtt{MIN} & \left[\mathtt{FCOMPL}\ -\right] \end{array}\right] \\
\mathtt{VAL} & \left[\begin{array}{ll} \mathtt{V\text{-}COMP} & \langle\,\rangle \\ \mathtt{SUBJ} & \boxed{2} \\ \mathtt{COMPS} & \boxed{3} \end{array}\right] \\
\mathtt{LEX} & +
\end{array}\right]
\end{array}\right] \right\rangle \\
\mathtt{SUBJ} & \boxed{2} \\
\mathtt{COMPS} & \boxed{3}
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

complements have to be consumed by the empty head before it combines with the finite verb.

(348) *head-vcomp-schema*

$$
\left[\begin{array}{ll}
\mathtt{SYNSEM} & \left[\mathtt{LOC}\ \left[\begin{array}{ll}
\mathtt{HEAD} & \boxed{1} \\
\mathtt{LEX} & - \\
\mathtt{VAL} & \left[\begin{array}{ll} \mathtt{VCOMP} & \langle\,\rangle \\ \mathtt{SUBJ} & \boxed{3} \\ \mathtt{COMPS} & \boxed{4} \end{array}\right]
\end{array}\right]\right] \\
\mathtt{DTRS} & \left[\begin{array}{ll}
\mathtt{HEAD\text{-}DTR} & \left[\mathtt{SYNSEM}\ \left[\mathtt{LOC}\ \left[\begin{array}{ll}
\mathtt{HEAD} & \boxed{1} \\
\mathtt{VAL} & \left[\begin{array}{ll} \mathtt{VCOMP} & \left\langle \boxed{2} \right\rangle \\ \mathtt{SUBJ} & \boxed{3}\langle\,\rangle \\ \mathtt{COMPS} & \boxed{4}\langle\,\rangle \end{array}\right]
\end{array}\right]\right]\right] \\
\mathtt{VCOMP\text{-}DTR} & \left[\mathtt{SYNSEM}\ \boxed{2}\right]
\end{array}\right]
\end{array}\right]
$$

Setting the values of the valence features can of course also be seen as a matter of parametrization to language specific values. For example, if one envisages to base-generate the subject in the *Vorfeld* or topic position in German, the SUBJ valence feature would be simply left unspecified in this rule schema. Likewise, it appears that in some free word order languages, such as Czech, the finite verb is not at all bound to a specific position, but may appear basically anywhere between the complements and the subject.[21] Again, this almost complete relaxation of word order constraints could be achieved by assigning unspecified values to all other valence features in this schema.

Of course, it could be argued that such an extension of the set of valence features implies exactly the kind of modification and extension of general principles and rule schemata that we argued against in the preceding section. This is true in the sense that in this version of our approach an additional carrier feature is introduced to transport the information of the initial finite verb, rather than exploiting existing ones. However, the proposed extension does not affect other parts of the Valence Principle. It is still fully consistent with the assumption that the relation between the empty clause-final node and the verb should be constrained to head domains. It does not lead to a duplication of the number of lexical entries required, and it allows one to maintain all the relevant generalizations on linear precedence that we introduced above.

5.3.4 V-initial and Separable Verb Prefixes

The approach of treating the finite verb in a clause-initial position as a lexical complement to a functional clause-final head has an almost obvious extension to the treatment of verbs with separable prefixes in German. The characteristics of these verbs are that they can form a discontinuous entity in the case of V-initial structures, where the stem occurs in a clause-initial position and the prefix is left behind in a clause-final position (349). In V-final structures the two substrings form a single word, as far as most orthographic representations are concerned (350).

(349) *Stellt* er das Problem *dar*

(350) daß er das Problem *darstellt*

In the following section we will first discuss different classes of data which serve to illustrate the problem and some possible solutions. We then show how and to what extend our approach carries over to the analysis of separable verb prefixes.

Some basic data and issues

The question of the relation between separable verb prefixes and the corresponding finite verb stem has been a matter of much discussion, which to date has not led

[21] Cf. for example [Sgall et al. 1973] and our discussion of [Oliva 1992] above.

to a fully convincing and satisfying solution. One of the reasons for this situation could be that typically a very broad range of phenomena at different levels of linguistic abstraction interact in this relation, so that it is hard to develop a solution which adequately incorporates the relevant phonological, morphological, syntactic and semantic constraints for all different variations of the construction. A second reason is presumably that, given that the relation between the two constituents is one of functor and argument or selector and selected, there are again no hard and reliable arguments for determining which one is which. We will mainly concentrate on this latter issue in the following and try to identify those cases for which it appears reasonable to treat the prefix as a functor.

It is quite obvious that the relationship between the prefix and the verb can be described along different parameters, not all of which appear to be discrete but could also be represented by a continuum. In particular, it is not always clear when and for what constructions one should postulate a transition from an argument of the verb, which can be interpreted on a compositional basis, to a semantically meaningless prefix, which could also be interpreted as a functor, and what criteria for identifying such a flip are.

Thus, one question that could play a role in this context is whether the meaning of the entire word form can be derived compositionally from the meaning of the prefix and the stem. Structures for which this appears to be case are on the one hand those constructions which are very close to support verb constructions or cases of object incorporation, such as *radfahren*, *skilaufen* etc.

Another group, in which some compositionality appears to be given, but where its full extent is not quite clear, are prefix verbs involving mainly some directional semantics. For these cases a sub-regularity seems to exist, that next to the simplex form (351) there is also a compound form with a non-separable prefix (352), as well as a separable prefix form (353), which all share the same basic semantic components.[22]

(351) Er *fährt* durch den Park

(352) Er *fährt* durch den Park *durch*

(353) Er *durchfährt* den Park

Apparently, the prefix versions, be they separable or not, are diachronically derived from a simplex form, either through incorporation or through duplication of the preposition. Yet, even if it is undebated that verb prefixes developed out of prepositions, it is not clear, which of the two latter versions should be seen as a precedent of the other. I.e., was the preposition first incorporated into the verb (352) and then duplicated and repeated on the direct object (353), or was

[22]For an extensive discussion of these constructions cf. [Wunderlich 1983].

there first a duplication of the preposition followed by an incorporation and 'transitivization' process?

For another class of constructions following the verb-prefix structure, a small clause or predication analysis appears to be the most appropriate. Again, there is a gradual transition from fully productive to more or less incorporated versions.

(354) Er schlug das Ei entzwei.

(355) Er tanzte sich die Füße wund.

(356) Er schrie ihn zur Türe hinaus.

Most of these constructions are closely related to the typical causative or resultative construction patterns in German. In particular, the constructions (355)/(356) are fully regular ad hoc formations and could be also analyzed as the result of a transitivization process involving resultative predication.

However, the question is whether we should let us deter by the apparent compositionality and productivity of these constructions from making the assumption that the verb should function as an argument rather than a functor in these cases. At least from a semantic point of view, it appears that the prefix could be considered equally well as a functor mapping the verb onto a more complex predicate. In the case of support verb constructions, for example, it has been quite often suggested that the non-verbal constituent carries most of the semantic content, while the verb only contributes some kind of aspectual information. Analyzing the verb as an argument rather than a functor is not so far-fetched then. Thus, while the plausibility of analyzing the prefix or prefix-like constituent as the functor in some or all of these constructions may vary, there is a priori no principled reason against treating prefixes as arguments.

One of the most interesting arguments *in favour* of this assumption comes from the area of coordination again.[23] Compare the following constructions involving coordination of some kind:

(357) Er macht die Türe auf und das Fenster zu

(358) *Er macht die Türe und sie haut das Fenster zu

(359) Er hat die Türe auf- und sie hat das Fenster zugemacht

Example (357) shows that we can easily have two different prefixes going together with the same verb and producing two different readings (*aufmachen* and *zumachen*). The admissibility of such a construction could be expected, if the

[23] The basics of this argument were first brought to our attention by Ulrich Block, already several years ago. See also [Netter and Wedekind 1986] for an LFG-analysis building on this idea.

prefixes were the functors which select a verb *machen* as a complement, and be it only on a phonological or morphological level. In (358) the reverse is tested, i.e., whether one prefix can satisfy the selectional requirements of two different verbs. However, the most natural reading for this sentence is a coordination of *machen* and *zuhauen*. This would be quite unexpected if the verb were the functor and the prefixes acted as arguments, since normally right node raising (or right periphery deletion in the first conjunct) of an argument is no problem whatsoever. In fact, the assumption that it is the prefixes rather than the verb which determine the semantics is even further corroborated by sentence (359), in which we find a kind or right node raising of the verb (similar to the construction in (358)), satisfying again the selectional restrictions of two different prefixes.

One could argue that this difference in grammaticality is due to the fact that the prefixes *auf* and *zu* are not entirely meaningless in these examples, but rather form a semantic opposition which is at the basis of the grammaticality of the contrastive constructions (357) and (359). While it cannot be debated that semantic contrasts do play a role in such coordination constructions, it is interesting to see that the basic distribution also persists with completely meaningless prefixes and even nonsensical coordinations. Thus, although the examples (360) and (362) sound quite funny and most likely will be interpreted as a pun, it is still not completely ungrammatical. In contrast, example (361) is again either ungrammatical or at least cannot be understood as a coordination of *anhören* und *anlachen*.

(360) Er hörte sich die Platte an und mit dem Lachen auf.

(361) *Er hörte die Platte und sie lachte ihn an.

(362) Er hat sich die Platte an- und mit dem Lachen aufgehört.

There are nevertheless some caveats which should be kept in mind when interpreting these data. Thus, the respective first construction in these sets of examples are most likely cases of verb gapping or at least can be closely related to such constructions (363).

(363) Er hörte sich die Platte an und sie mit dem Lachen auf.

For gapping, however, we know that it is always the topmost finite verb plus any unbroken and direct chain of functors or heads depending on this finite verb together with their dependent arguments which can be elided in the second conjunct. In example (364) neither of the two verbs in parenthesis, which would break this chain of functors, can be present without making the sentence ungrammatical.

(364) Er soll neulich versucht haben, mit dem Rauchen von Marihuana aufzuhören und sie jetzt (*aufgegeben) von Pampagras (*anzufangen)

Thus, if the prefix constructions (360)/(363) are also gapping constructions, the gapping chain must indeed begin with the topmost finite verb rather than the topmost functor, since otherwise the prefix would have to be elided in the second conjunct too.

Secondly, there may well be an interference of the prosodic structure in the cases of right node raising. Since it is always the prefix which carries the main word stress, a coordination construction which requires the stressing of the prosodically subordinate verb (i.e., (357) and (360)) could be less natural than one in which the prefixes carry the main stress and thus comply with prosodic constraints at both the word and the sentence level. On the other hand, the fact *that* the prefix carries the main word stress could also be interpreted as an indicator that this element is the appropriate functor, carrying the relevant (semantic and syntactic) information.

Being well aware of all these possible imponderabilities, we will nevertheless explore in the following how an analysis of separable verb prefixes as functors or heads will fit into our approach to functional heads in general and our analysis of V-initial in particular.

Separable verb prefixes as functional heads

Since we do not assume a movement or reconstruction analysis, through which the finite verb is reinserted in a V-final position at run-time, we have to assume that the verb stem and the prefix are produced from the full composite form by means of a lexical rule. Although in vast majority of cases a homonym of the stem or the prefix will of course already exist in the lexicon, there is still a small number of lexical items such as *ab-kanzeln* or *dar-legen* for which either the stem or the prefix do not have the status of an independent morpheme. If we then want to treat the prefix as a (non-empty) functional head in V-initial constructions analogous to the empty functional head occuring with simplex verbs, we have different choices for splitting the information of the original lexical item between the stem and the prefix.

One option is to make the type for the prefixes almost completely identical to the empty heads, such that the full original valence information as well as the semantic information all goes to the verb stem and is then adopted from the prefix selecting this verb as a complement. The biggest disadvantage of this approach is that the prefixless finite verb under this assumption has exactly the same structure as the one with the prefix. As a consequence one would need some additional feature or mechanism for blocking the derived prefixless form from occuring in a V-final position on its own.

As an alternative, the valence features of the original as well as its semantic structure could go to the prefix, which also has to subcategorize for the verb stem, of course. In contrast to other functional categories, the prefix would thus carry substantial syntactic and semantic information on its own, while it shares all the other properties of functional heads, viz., the co-indexation of the major features, the triggering of functional completeness, etc. The feature structure in (365) illustrates the representation of a prefix as a functional head along these lines.

(365) Separable Prefix as Functional Head

$$
{}_{\textit{prefix}}\begin{bmatrix}
\mathtt{LOC|CAT} & \begin{bmatrix}
\mathtt{HEAD} & \begin{bmatrix}
\mathtt{MAJ}\ \boxed{1}\begin{bmatrix} \mathtt{POS}\begin{bmatrix}\mathtt{N} & - \\ \mathtt{V} & +\end{bmatrix} \\ \mathtt{AGR}\ \boxed{2} \\ \dots \end{bmatrix} \\
\mathtt{MIN}\begin{bmatrix}\mathtt{FCOMPL} & + \\ \mathtt{SPEC} & -\end{bmatrix}
\end{bmatrix} \\
\mathtt{VAL} & \begin{bmatrix}
\mathtt{VCOMP} & \left\langle {}_{\textit{vfin}}\begin{bmatrix} \mathtt{LOC} & \begin{bmatrix} \mathtt{CAT} & \begin{bmatrix} \mathtt{HEAD} & \begin{bmatrix}\mathtt{MAJ} & \boxed{1} \\ \mathtt{MIN} & [\mathtt{FCOMPL}\ -]\end{bmatrix} \\ \mathtt{VAL} & [\,] \\ \mathtt{LEX} & + \end{bmatrix} \\ \mathtt{MORPH} & \begin{bmatrix}\mathtt{HEAD} & [\,] \\ \mathtt{STEM} & \textit{string}\end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\
\mathtt{SUBJ} & \left\langle [\mathtt{LOC|CAT|HEAD|MAJ|AGR}\ \boxed{2}] \right\rangle \\
\mathtt{COMPS} & \langle .. \rangle
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

In this structure there is no relation between the valence features of the finite verb and the valence of the prefix, since the latter is assumed to be derived from the original prefix verb. Since the prefix has full access to the agreement features of the finite verb through co-indexation of the MAJOR feature, the subject-verb agreement can be triggered by the prefix. The prefix of course must not carry a [SPEC +] specification, since then both adjunction to the prefix would be blocked and the linearization rules would yield the wrong results.

One major change which we assume relative to standard HPSG is the introduction of an additional layer of representation for the morphological structure of strings encoded in a feature MORPH, which occurs next to the syntax (CAT) and the semantics (CONTENT) of an expression. In this level of abstraction we assume the morphological structure of words to be represented, containing at least a STEM attribute which abstracts from inflectional variations of the word forms, together with a morphological HEAD feature which encodes the inflectional properties of a word. In principle, the verb can be selected by the prefix by constraining this level of representation only, since the only information required is the stem of the verb and its morpho-syntactic inflectional properties. This means that we do not have to generate a second entry for verbs which already exist as simplex forms, but that we can make use of the existing forms as complements of the prefix and selectively only access their MORPH feature while ignoring all the other syntactic and semantic properties.

Note by the way that the introduction of such an additional level of representation will be inevitable anyway, if HPSG wants to come to a more comprehensive treatment of idiomatic or semi-idiomatic expressions. In many cases, idiomatic expressions do not consist of fixed phonological or orthographic strings, but involve a certain amount of variation over inflectional properties, for example by active/passive alternations. For these cases, the abstractions of a morphological structure could be the appropriate level of representation.[24]

5.3.5 Summary

In this section we showed how our analysis of functional heads can be extended to the treatment of finite verb position in German, involving both overt functional categories, such as complementizers and separable verb prefixes, and empty functional heads in the case of V-initial structures.

While the former class of lexicalized functional categories followed the generalizations of the previous chapter quite smoothly, we had to make some additional assumptions for the empty functional head triggering structurally induced functional completeness. We nevertheless showed that this approach at the same time can explain the complementary distribution of V-initial structures and complementizers, while it does not require any fundamental changes of the principles of HPSG but also avoids the redundancies caused by lexical rules duplicating the lexical entries for finite verbs. Although the extension of the approach to separable prefixes is clearly pending on some empirical questions concerning the different status of the prefixes, we showed that it is technically feasible and that it fits into the framework quite naturally.

[24]Cf. [Erbach and Krenn 1994] who also envisage such a level of representation as an alternative to a LEXEME feature roughly corresponding to the STEM attribute above.

5.4 Relative Clauses

We conclude this chapter by discussing the phenomenon of relative clauses. We are confident that we can present a better solution for some aspects of these constructions than it is suggested in standard HPSG. However, relative clauses also stand for a class of constructions which represent an open problem for our approach of functional heads.

Relative Clauses so far did not have a satisfactory solution in HPSG, since their treatment again involves the introduction of an empty head, for which it is unclear whether it is fully indispensable. Quite recently a revision of this approach was suggested by [Sag 1995] under which the verbal head is treated as the head of the clause. Below, we will suggest another alternative treatment, which is based on an extension of the set of selector features.

A treatment of relative clauses has to capture at least the following aspects:[25]

(a) *Topicalization*: Relative clauses (in Germanic languages) involve a non-local dependency in the form of left dislocation or topicalization of a phrase (forthwith WH-phrase) containing a relative pronoun.

(b) *Pied Piping*: The relative pronoun can be embedded in the WH-Phrase at various depths. While English permits almost arbitrary embedding, the relative pronouns in German can be maximally embedded under a PP node and in a prenominal genitive.[26]

(366) der Mann, mit dessen Vaters Bruder ...

(c) *Linear Precedence*: Independent of the specific solution to Pied Piping, German has a strong constraint on linear order to the effect that a relative pronoun must be the head of the first nominal constituent in a relative clause.

[25]We ignore reduced relative clauses in English and other languages, and concentrate on the description of German here.

[26]Whether further embedding under some verbal projection has to be envisaged for German, as for example in (i), is unclear.

(i) ein Mann, mit dem sich zu streiten keine Freude macht

For these constructions there are two competing analyses, one which embeds the WH-Phrase further under a VP [van Riemsdijk 1985], another which analyzes the string *mit dem sich zu streiten* as two separate constituents, which are more or less accidentally adjacent to each other, but could also be separated [Haider 1985].

(ii) ein Mann, [mit dem] wahrscheinlich [sich zu streiten] keine Freude macht

(d) *Adjunction*: With the exception of free relatives which have to obey quite different conditions, relative clauses have to be combined with a reference noun. This combination is quite similar to an adjunction relation, since relative clauses are optional and there may be more than one relative clause modifying a head noun.

(e) *Agreement*: In some languages, as for example German, relative clauses have to show (syntactic) agreement with the reference noun with respect to the features `NUMBER` and `GENDER`.

(f) *Head Position*: Relative clauses are consistently V-final in German.

In the following we will first present the standard HPSG approaches and our own alternative variant, which offers a solution to the issues (a) to (e). We will then discuss the problem for our analysis of functional heads which arises from (f) and which is mainly caused by the fact that HPSG (other than e.g., Categorial Grammar or Tree Adjoining Grammar) does not foresee the notion of a phrasal functor whose functor properties are not anchored in the lexical head of the phrase.

5.4.1 Standard Approaches in HPSG

To capture the above-mentioned generalizations [Pollard and Sag 1994] suggest the introduction of an empty head through which most of the constraints are mediated. This empty head (367) takes two complements as arguments on its `SUBCAT` list; the first one is the body of the clause with the missing WH-phrase which occurs to the right of the empty relativizer, the second one, to the left, is the WH-phrase itself. Through the lexical entry the `LOCAL` value $\boxed{2}$ of the WH-phrase is then identified with the `SLASH` value of the clausal complement.[27]

(367)

[27]Note by the way that the complementational aspects of this approach would be difficult to translate into the revised version building on a split between the `SUBJ` and `COMPS` selection. Under the assumption that the WH-Phrase is the only complement of a phrasal head occuring to its left, it more or less automatically acquires the status of a subject, which does not really contribute to a more homogenous notion of subject. Alternatively one would have to assume a flat structure similar to the auxiliary inversion constructions, in which the empty head appears to the left of both of its complements, which then have to be ordered by means of an LP constraint.

$$\begin{bmatrix} \text{LOC|CAT} & \begin{bmatrix} \text{HEAD|MOD} & \begin{bmatrix} \text{TO-BIND|REL} & \left\{\boxed{1}\right\} \end{bmatrix} \\ \text{SUBCAT} & \left\langle \begin{bmatrix} \text{LOC} & \boxed{2} \\ \text{N-L} & \begin{bmatrix} \text{INH|REL} & \left\{\boxed{1}\right\} \end{bmatrix} \end{bmatrix}, {}_{\mathit{sfin}}\begin{bmatrix} \text{N-L|INH|SLASH} & \left\{\boxed{2}\right\} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{NON-LOC} & \begin{bmatrix} \text{TO-BIND|SLASH} & \left\{\boxed{2}\right\} \end{bmatrix} \end{bmatrix}$$

However, the function of the empty head is not only to establish the relation between the WH-phrase and the body of the relative clause, but also to relate the entire relative clause to the reference noun modified by it. Therefore the empty head in addition has a `MOD` feature through which the noun is selected. Agreement between the noun and the relative pronoun is achieved through a non-local `REL` feature which percolates within the WH-phrase from the relative pronoun to the maximal projection and which contains the relevant agreement information, in terms of a semantic index. This index $\boxed{1}$ is identified, again through the empty head, with the index of the modified noun through the empty head's `MOD` feature.

In order to satisfy the Non-Local Principle ([Pollard and Sag 1994][184]), repeated in (368), and to prevent a further projection of the `SLASH` feature outside the relative clause, the empty relativizer introduces a `TO-BIND|SLASH` feature, which is cancelled as soon as the relativizer combines with the finite clause.

(368) **Non-Local Principle** For each non-local feature the `INHERITED` value on the mother is the union of the `INHERITED` values of the daughters minus the `TO-BIND` value on the head daughter.

Secondly, to prevent projection of the `REL` feature beyond the attachment node of the relative clause to the noun, the empty head imposes a constraint on the reference noun that it must have a `TO-BIND|REL` value which is identical to the `INHERITED|REL` feature of the WH-phrase. This latter constraint introduces a certain twist to the definition of nouns, since there are basically only two possibilities how the lexical entry of a (reference) noun can be specified. Either the value of its `REL|TO-BIND` feature is defined as a disjunction between an empty set and a non-empty set, or it has $\top$ (or rather some type defining a set with an open arity) as a value. If the former variant is chosen much of the elegance of this solution is lost, since one could have just as well a disjunction specification on the noun through which it optionally selects the relative clause. On the other hand, if the latter option were chosen, the union of the `INHERITED` values can never be properly computed, since it would contain an unconstrained and underspecified variable (the `REL|TO-BIND` value) which still remains to be subtracted.

As an alternative to this approach, [Sag 1995] suggests to treat the topmost verbal head of the relative clause as the head of the entire relative clause. Except for some small differences this verbal head takes over the functions of the empty relativizer.

However, it also means that for every verb two different lexical entries have to be envisaged, which can be related by a lexical rule and function as normal head in all non-relative clauses and as enriched heads taking over the relativizer function in relative clauses.

We do not challenge the technical feasibility of these approaches, but rather want to argue that they lack a certain elegance and intuitiveness, even if the second approach presents a remarkable advance over the first one. The main criticism of the analysis involving an empty relativizer concerns not only the fact that it relies on the introduction of an empty head, but also and above all the overloading of the functionality of this empty relativizer which has to account for a whole series of relations which intuitively should be attributed to other parts of the construction. For example, what appears to be a standard non-local dependency throughout, viz., the dislocation of the WH-phrase, is translated by the empty relativizer from a head-complement relation into a non-local relation. Secondly, although the relative pronoun is the element which both licences the clause to function as a relative clause and carries the agreement and co-reference properties with respect to the reference noun, its role is completely downgraded to an element whose presence has to be detected and mediated by other elements. In particular the last aspect also holds for the new approach by [Sag 1995], which in addition also suffers from the redundancy caused by the lexical duplication of the verbal head.

Still, before we develop an alternative analysis below, we want to emphasize that the approaches described above could easily be accommodated with our analysis of functional heads. For the first approach this would simply require that we treat the empty relativizer as a functional head subcategorizing for the two complements in the way described above. One of these complements functions as the functionally incomplete, viz., the verbal projection or body of the relative clause, the other one is an additional argument represented by the WH-phrase. In a sense, such an analysis could be seen as another instance of functional completeness being brought about by structural means similar to our treatment of V-initial.

In the case of the verbal head taking over the functionality of the empty relativizer, an incorporation into our approach would be equally trivial. The lexical rule relating the two verbal entries simply has to switch the marking of the finite verb from functionally incomplete to complete. As a result we obtain a finite verb which cannot occur in a V-initial position, since the empty functional head only accepts a functionally incomplete clause, and which for the same reason cannot occur in a V-final position in combination with a complementizer. The arguments we levelled against these two approaches would keep their validity, however. We will therefore explore an alternative, which unfortunately suffers from the disadvantage, however, that it can be less easily integrated with our analysis of functional heads.

5.4.2 REL as a Selector Feature

In the following, we explore an alternative version of the treatment of relative clauses, which requires some extensions of various HPSG assumptions in terms of introducing an additional selector feature and an additional schema, but which has the advantage of dispensing with the empty relativizer and re-allocating its functionality to the relative pronoun in an intuitively more appropriate way.

As a presupposition of the elimination of the empty relativizer, we have to give an alternative account for its three main functions, viz., (a) the mediation between the dislocated WH-phrase and the body of the relative clause, (b) the selection of the reference noun and the establishment of an agreement relation and (c) the integration of the relative clause's semantic structure with the semantic structure of the reference noun. We intend to cover (a) by the standard head-filler-schema, (b) by introducing a selector feature REL which is lexically anchored in the relative pronoun, and (c) by introducing an additional rule schema.[28]

In the standard approach the combination of the relative clause with the reference noun is achieved by a head-adjunct-schema, i.e., a schema which combines a non-head functioning as a selector with a head functioning as an argument. Since MOD is a head feature, the empty relativizer has to be assigned the function of a head in order to licence the schema. In this sense, the empty head is simply some means to hide the fact that there is no overt head in a relative clause which could appropriately function as the lexical basis for the selector feature necessary for the selection of the reference noun in the head-adjunct-schema.

Now, instead of covering up the lack of an appropriate head in this sense by either introducing an empty head (or by making the verb the relevant head), we suggest to change the selection mechanism and to allow a non-head to be the basis for the selector feature. Our alternative is based on the idea that the attribute REL, which in the standard approach serves as an non-local carrier of the agreement information, is treated as the primary selector feature which turns the relative clause into a functor and which selects the reference noun as an argument. As a selector feature, REL will thus have a similar status as COMPS and SUBJ, MOD and also SLASH. We assume that this feature will be introduced or lexically anchored in the relative pronoun, which can either be equivalent to the WH-phrase itself or it can be embedded into this phrase. Since in the latter case the pronoun need not function as the head of the WH-phrase and since in any case the entire WH-phrase does not fulfill the function of a head in the relative clause, the REL

[28] Note that in the approach by [Sag 1995], the aspect of (a) is likewise re-allocated to the head-filler-schema, (b) is only partly changed, as selection is still assigned to the MOD feature and is simply shifted from the empty relativizer to the verb, and (c) is fulfilled by the standard head-adjunct-schema. Although this looks like a much more conservative extension at first sight, it should be kept in mind that it has to be paid for by duplicating the lexical entries of the verbs.

feature will have to share the property of the SLASH feature that its percolation is not constrained to a spine of HEAD-DTRs.

This means that we have to extend the set of possible grammatical relations by another relation or rule schema in which a non-head as a functor selects a head as an argument, similar to the head-adjunct-schema, which is different however in that the functor property or selector feature is not projected from the lexical head of the non-head functor, but may come from a non-head. This schema, which we call the head-rel-schema fills a logical gap in a characterization of possible grammatical relations which is based on the combination of the three parameters (a) the function of the lexical anchor of the selector feature, (b) the local function of the functor or selector and (c) the local function of the argument. The existing schemata and the head-rel-schema will fit into the following matrix:

Schema	Lexical Root	Local Functor	Local Argument
head-comps	head	head	non-head
head-adj	head	non-head	head
head-filler	non-head	head	non-head
head-rel	non-head	non-head	head

The fact that not all possible combinations in this matrix are exploited could be attributed to the assumption that the local functor and argument cannot both function as heads or non-heads simultaneously within the same rule. Thus, in some respect the third parameter is practically redundant, since it only reverses the second one. It could become relevant, however, if we foresee a constellation in which all three parameters are set to the same value “head” (or “non-head”). A possible instance of this configuration could be a rule schema for coordination.

The technical details of our approach are then as follows. The WH-phrase is treated as a standard FILLER-DTR and its non-local dislocation is accounted for by the normal SLASH mechanism. To ensure that the occurence of relative pronouns is restricted to the domain of the FILLER-DTR or WH-phrase, we extend the filler-head schema such that the HEAD-DTR is required to have an empty-list-valued REL feature. This extension slightly deviates from the specification of the schema suggested by [Pollard and Sag 1994], but does not really affect the theory since it can be realized by adding a simple constraint to the filler-head schema.[29]

The relative pronoun introduces the selector feature REL in its lexical entry. The value of this feature can be specified as a singleton list whose element is constrained

[29]Note that [Pollard and Sag 1994][216] propose to treat clauses (as opposed to VPs) as barriers for relative pronoun extraction. Depending on how clausal projections are defined relative to the filler-head constructions, this assumption could be consistent with the above.

to the type of the SYNSEM value of the possible reference nouns. Since the feature must percolate along paths other than projections of the head, we declare it to be a LOCAL feature, analogous to the valence features.

To discharge the selection of the reference noun, we postulate a head-rel-schema which requires the head to be a noun or nominal projection and the functor-daughter to be a sentence with a non-empty REL-feature. The value of the REL feature is identified with the SYNSEM value of this nominal head, along the lines of complement selection and consumption.

The definition of the percolation of the REL feature from its lexical anchor follows the lines of the other selection principles, which are also expressed as default principles:

(369) **REL-selection Principle** In a head-rel-schema the REL value of the mother is the value of the REL value of the REL-DTR minus the SYNSEM value of the HEAD-DTR. In all other cases the REL value of the mother is the concatenation of the REL values of the daughters.

While list operations can be relatively expensive, we can implement this principle and the respective constraints by means of much more efficient difference lists. To ensure that only one relative pronoun occurs in a clause, we specify that the lexical entries of all relative pronouns have a REL feature whose value is a list of length one (370) (a), i.e., the value of LAST.REST is set to $*end*$ instead of an open variable. The REL value of all other categories is specified as an empty difference list (370) (b).

$$(370)\quad (a)\ \begin{bmatrix} \text{LIST} \begin{bmatrix} \text{FIRST} & noun[\ ] \\ \text{REST} & \boxed{1} * end* \end{bmatrix} \\ \text{LAST}\ \boxed{1} \end{bmatrix} \qquad (b)\ \begin{bmatrix} \text{LIST}\ \boxed{1} \\ \text{LAST}\ \boxed{1} \end{bmatrix}$$

The head-rel-schema roughly has the form given in (371). The specification of the difference list subtraction in this schema is special insofar as the LAST value of the REL-DTR is not co-indexed with the mother's LAST value, which allows for the possibility that relative clauses can be embedded into each other. For all other schemata the standard difference list concatenation is assumed.[30]

[30]To determine the specific linear constraints within the WH-phrase a combination of structural and lexical constraints of empty difference lists can be envisaged, through which the "forbidden" domains for relative pronouns are closed. For English, the above-mentioned limitation of relative pronouns to the WH-phrase will already do the bulk of the work.

(371)
$$\begin{bmatrix} \mathtt{SYNSEM} & \begin{bmatrix} \mathtt{LOC} & \begin{bmatrix} \mathtt{CAT|REL} & \begin{bmatrix} \mathtt{LIST} & \boxed{1} \\ \mathtt{LAST} & \boxed{1} \end{bmatrix} \\ \mathtt{CONT} & \begin{bmatrix} \mathtt{RESTR} & \boxed{3} \bullet \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \mathtt{DTRS} & \begin{bmatrix} \mathtt{HEAD\text{-}DTR} & \begin{bmatrix} \mathtt{SYNSEM} & \boxed{2} \begin{bmatrix} \mathtt{LOC|CONT|RESTR} & \boxed{3} \end{bmatrix} \end{bmatrix} \\ \mathtt{REL\text{-}DTR} & \begin{bmatrix} \mathtt{SYNSEM|LOC} & \begin{bmatrix} \mathtt{CAT|REL} & \begin{bmatrix} \mathtt{LIST} & \begin{bmatrix} \mathtt{FIRST} & \boxed{2} \end{bmatrix} \\ \mathtt{LAST} & *end* \end{bmatrix} \\ \mathtt{CONT|RESTR} & \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

In this schema, the combination of the relative clause's semantics with the semantics of the reference noun is sketched by a simple concatenation of the semantic constraints of the two daughters. This stands in contrast to head-adjunct constructions, where the standard Semantics Principle specifies that the lexical head of the adjunct is responsible for the integration of the semantics of the adjunct and the head. Since in the case of the head-rel-schema the selector is not identical to the lexical head of the non-head daughter, the Semantics Principle cannot be applied in its standard way, but has to be replaced by a structural or rule-based combination of the semantics in this schema.[31]

5.4.3 Open Issues

If one adopts the above approach to the treatment of relative clauses, some further issues arise in connection to our treatment of functional categories. At this point, we can only present the problem and sketch some tentative solutions, but we will have to refer a substantial part of the problem to future research.

Functional completeness in relative clauses

One of the most fundamental restrictions of our analysis of functional heads is that we have to assume that the property of functional completeness is projected from a lexical anchor which also functions as the head of the maximal projection. In our relative clause analysis above we had to face a similar problem in the form of the selection of the reference noun and circumvented it by introducing

[31] In the implementation of this fragment, for which the semantic specifications were made by Walter Kasper, a structural combination of the semantics is employed throughout the grammar also for standard head-adjunct constructions.

a percolation mechanism for selector features which is not bound to head paths. This solution cannot be applied to functional completeness, since this property is inherently bound to the `HEAD` features.

Under the above `REL` analysis, however, we have to cope with the conflict that the verb functions as the head of a relative clause and that finite verbs by definition are functionally incomplete. As we already mentioned, this problem would have a very simple solution under the approach involving an empty relativizer along the lines suggested by [Pollard and Sag 1994]. On the other hand, if we were willing to change the assumption about the functional completeness of the finite verb or even admit a second entry which is functionally complete, we could just as well adopt the analysis of [Sag 1995] and treat the verb as the lexical basis of the functor properties of the relative clause.

Thus, the only remaining directions to look for are either that we can determine some other constituent which could plausibly function as a head of the clause, or that we give up the assumption that all clauses must be marked as functionally complete.

The only likely candidate for an alternative head, along the lines of the first option, would be the entire WH-phrase in the topicalized position. This means that the analysis of this configuration as an instance of the filler-head schema would have to be revised and be substituted by an analysis in which the WH-phrase functions as a *phrasal* functional head taking the body of the clause as a complement.

The crucial problem that we encounter under this assumption is of course that phrasal heads are always defined in terms of the lexical head from which they are projected. Again, the prime candidates for such lexical heads would be the class of relative pronouns. However, even in German it is more or less accidental that the relative pronoun functions in the majority of cases as the head of the WH-phrase. At least all more complex pied piping constructions involving a WH-phrase in the form of prepositional phrase would not be covered by this assumption and would require some other solution.

The second option is to turn the problem into an asset and to claim that relative clauses differ from other types of clauses exactly in being inherently functionally incomplete. The type definition for relative clauses would then require them both to have a non-empty `REL` attribute and to be functionally incomplete. This would be an advantage insofar as we could claim that no relative clause all by itself could function as a root constituent but first has to combine with some reference noun. As it is well known, most other clausal constructions occuring in a subordinate function can also be used as independent root sentences, typically with some quite specific speech act interpretation.[32]

[32]Cf. for example [Weuster 1983] for an overview.

The main problems that could arise from such an assumption is the close parallelism between relative clauses and indirect interrogatives involving a WH-phrase containing some interrogative pronoun.

(372) Ich will wissen, [mit wessen Bruders Vater] du gekommen bist.

If we assume that relative clauses are functionally incomplete, it is hard to see how we could classify WH-clauses such as (372) differently. The inconsistencies arising from such an assumption are twofold; (a) V-final interrogative WH-clauses can very well function as independent root sentences, e.g., as deliberative questions with the modal particle *wohl* and (b) subordinate interrogatives with the conjunction *ob* would have a different status than the WH-interrogatives and we would thus either need some disjunctive subcategorization of the matrix predicate or some generalization over the two different types.

In short, whatever solution we come up with to enforce a V-final position in WH-interrogatives and relative clauses and whichever way we decide on the issue of functional completeness, the result has to be consistent with a more general theory of clause types. Before we turn to this issue, we want to point out once more that the problem we are facing here is not necessarily exclusively due to our particular theory of functional heads, but may well be a more general problem of the relation between the notions of head, functor, selector and lexical anchor in HPSG.

Clause Types

One aspect of the issue we just discussed goes back to the question whether all maximal projections should be defined as functionally complete. Conversely we could ask what the exact definition of a maximal projection should be or even what the purpose of a maximal projection should be. The answers to these questions were comparatively simple in the area of nominal categories, where we could use paradigmatic relations to characterize the maximal nominal projections. As we already pointed out even then, there are some cases, such as predicative nominal phrases in copula constructions, where functional completeness is not a required property.

In the verbal domain, there is an even greater heterogeneity of clausal and sentential structures that may count as a maximal projection, in particular if we take into account all the different types of subordinate and independent clauses. While we probably did well to concentrate on our main issue of functional categories which lead to a selection of a subset of these and not to let us being distracted by the possible definitions of clause types, we want to have a brief look at how and where our theory of functional categories could fit into a theory of clause types.

Such a theory of clause types is necessary above all for two reasons, (a) for the definition of selectional and distributionally relevant constraints in the case of subordinate clauses, and (b) for the definition of possible root sentences in terms of *sentence mood*, i.e., syntactic configurations which contribute to the determination of speech act interpretations.

Mainly to clarify the issue and to summarize the scope of our results, rather than to propose a thoroughly worked out theory of clause types, we list in the following table all those distinctions that we can easily capture with our approach to functional categories and finite verb position so far. The parameters or features that we assume are the two `MINOR` features `FCOMPL` and `SPEC`, a structurally introduced `LOCAL` feature `TOPIC`, and the lexically rooted `NON-LOCAL` feature `QUE`. The clause types distinguished are characterized by finite verb position, the subordinate vs. main clause distinction and some main sentence mood type.[33]

Clause Type	FCOMPL	SPEC	TOPIC	QUE
V-2 Declarative Main	+	-	+	-
V-F Declarative Sub	+	+	-	-
V-1 Interrogative Main	+	-	-	[ ]
V-F Interrogative Sub	+	+	-	+
V-2 WH-Interrogative Main	+	-	+	+
V-F WH-Interrogative Sub	-	-	+	+
V-F Relative Clause	-	-	+	-
V-F Adverbial Sub	+	+	-	-

For most of these features we already defined in detail how they can be introduced. The [`FCOMPL` +] specification, for example, is introduced either through the clause-final empty head or through a complementizer. As we just pointed out, the only cases where we so far did not find an appropriate way for introducing this feature are the V-final clauses with a WH-phrase in the topic position.

In a subset of the constructions marked as functionally complete, the specification `SPEC`+ is introduced by the complementizers as part of their `MINOR` feature. All other (empty or non-empty verbal) heads are specified accordingly as [`SPEC` −]. Since the `SPEC` feature in the verbal domain is bound to complementizers, one could even envisage as its possible values a more complex sortal hierarchy, rather than the simply binary distinction, which could serve to distinguish the different types of complementizers, viz., the complementation complementizers *daß* and *ob*, as well as the different adverbial complementizers *als*, *nachdem*, *obwohl*, etc.

[33] Quite obviously there has to be some redundancy in this table, since three binary features should be enough to distinguish eight items. However, we have to keep in mind, that not all combinations will yield natural configurations for clause type specifications, and that among the combinations which are possible the distinguishing information should be introduced in an adequate and systematic way by some (lexical or structural) device.

This would easily compensate the lack of a distinctive feature in the encoding of V-final adverbial and declarative clause, i.e., *daß* clauses.

The attribute TOPIC we implement as a LOCAL feature, which is introduced by structural rules. Specifically the filler-head-schema and equivalent rule schemata will be marked with a positive value for this feature at the mother node, all others have a negative value, which will allow us to block recursive applications of the filler-head-schema or applications in the clausal *Mittelfeld*.

The QUE specification is already foreseen in standard HPSG as a property of interrogative pronouns. It could be either treated as a semantic property or alternatively as a NON-LOCAL feature which is not projected across maximal projections except across the root node of the phrase in the topic position. The reason for this restriction is that only WH-pronouns in this position satisfy selectional constraints of governing items. Of course, WH-pronouns may also occur *in situ* , i.e., in positions other than the topic, which should trigger an interpretation as an echo-question. Co-occurence of WH-pronouns *in situ* in addition to a WH-pronoun in topic position should lead to an interpretation as a multiple WH-question.

Whether or not the complementizer *ob* should bear a QUE specification we can leave open here, since this clause type can be distinguished equally well by other features, in particular SPEC. Note in this context that Y/N-interrogative main clauses cannot be marked as [QUE +], while their subordinate equivalents could. The reason for this is again that in the former case there is no place where we could plausibly introduce this property (except for a root node), while there the complementizer provides a lexical anchor for this feature.

It should be clear that in this table we abstracted away from a large number of pragmatically relevant distinctions which are derivative or which are encoded by other properties anyway. For example, the distinction between imperatives and Y/N-interrogatives which are both subsumed under V-1 main clauses, will already be marked in the MAJOR features of the verbal HEAD. For those cases, where such a marking is missing, i.e., *Sie*-imperatives or adhortatives (*Gehen wir*), the ambiguity between imperative and interrogative interpretation has be resolved at a different level anyway.

Above all we did not provide any generalizations about the distribution of the different clause types in subordinate positions. Thus, most of the clause types which we declared to be main clauses, could also occur in a subordinate complement or adjunct position. Finding the natural classes and generalizations for these distributions is a rather complex enterprise and one of the main reasons why we declared this area to be an open issue for us.

5.4.4 Summary

In this section we explored a set of constructions which were not yet covered by our analysis of verbal functional categories. We first compared two approaches in standard HPSG, which build on the introduction of a specific lexical head for relative clauses, either in the form of an empty relativizer or in the form of a second lexical entry of the verbal head. For both of these approaches, it would have been absolutely trivial to integrate them with our theory of functional heads. However, since both of these approaches suffer from a certain amount of redundancy, we tried to explore an alternative in which the relative pronoun is treated as the lexical anchor of the relevant selectional properties of the relative clause.

Following this analysis, we also pushed our analysis of functional heads in the verbal domain to its limits, since under this assumption there is no lexical head from which the property of functional completeness can be appropriately projected if more complex cases of pied piping are also taken into account. We therefore raised the question whether functional completeness must be a property which has to hold for every clausal projection and identified relative clauses and V-final subordinate WH-interrogatives as the crucial and problematic cases, among the broad range of clause types covered by our approach.

5.5 Summary

In this chapter we explored how the notion of functional heads, as we developed it in the preceding chapter on the basis of nominal phrases, can be exploited and applied in the clausal domain. This transfer from one domain of categories to another presented both an opportunity and a challenge.

On the one hand, we were able tho show that a substantial part of the theory carries over with hardly any modifications whatsoever. This concerned above all the application of the notion of functional heads to constructions with complementizers, i.e., those categories which have been traditionally very close to the class of lexical functional categories.

On the other hand, we also had to face the problem that the full range of phenomena determining German clause structure presents a highly complex and interwoven set of issues, not all of which have found a standardized and unanimously accepted answer in the literature. In particular, the problem of the (clause-initial) position of the finite verb and of the comparatively free order of nominal and prepositional complements in the *Mittelfeld* has given rise to a broad range of different analyses with quite different implications. While it was interesting to see that our theory would be quite compatible with many of these analyses and in fact has been used as the basis for an integration of several of these different approaches, we nevertheless pursued our own alternative analysis of V-initial structures.

In this analysis, we were able to maintain several highly desirable properties, such as a uniform structure for V-initial and V-final clauses, the preservation of basic generalizations and principles of HPSG, as well as a natural account for the complementary distribution of complementizers and finite verbs in an initial position. The approach crucially builds on a different form of functional completeness, viz., a structurally induced functional completeness encoded by means of an empty functional head, rather than the lexically induced functional completeness that we encountered with lexical specifiers. As an extension to the approach, we showed how these basic assumptions also carry over to an analysis of separable verb prefixes in which the prefix is assigned the role of a functor which takes the verb stem as an argument.

In the final section, we completed the analysis of different clausal constructions by investigating different approaches to relative clauses. These constructions are representative for a class of configurations which evade a straightforward analysis on the basis of our theory of functional heads, unless an artificial functional head in the form of an empty relativizer or a duplicate verb entry is introduced. Nevertheless, the dilemma that manifests itself in these constructions is quite illustrative. Either one opts for a more redundant and less constrained analysis and can thus maintain certain generalizations, or one decides in favour of an optimally restricted methodology and may be forced to abandon or at least to extend certain other basic assumptions. In our case, the assumption at stake was the generalization that all maximal projections should be marked as functionally complete. Giving up this generalization could give us more leeway in dealing with the whole range of different clause types, maintaining it will make our theory stricter and more consistent, but it will also force us to fill the potential predictive gaps by some other means.

Chapter 6

Conclusion

With the work that we have presented here, we attempt to make a contribution to the computational linguistic description of German, as well as to the theory of Head-Driven Phrase Structure Grammar. On a more abstract level, some aspects of this work can also be viewed as a theoretical exercise in opening up new paths, which lead away from traditional phrase structure and $\overline{\mathrm{X}}$-oriented approaches.

As far as the description of German in a high-level grammar formalism, viz., typed feature logics, is concerned, we have covered a broad range of core data of German syntax in the areas of nominal and verbal projections. We tried to keep these descriptions as modular as possible, such that different parts of the analysis can also be used as results in isolation. However, we also attempted to make sure that the individual areas which were worked out in detailed descriptions fitted together on a higher level, either by making it compatible with general assumptions of HPSG or by sketching the possible and necessary links and extensions.

Of course, all grammars leak, and the fragment of German that we specified still leaves a number of open issues which require some further research, as we already pointed out. The example of relative clauses and subordinate WH-interrogatives is just representative for a general phenomenon that we also find in the domain of nominal constructions in the form of prenominal genitives, e.g., *Peters guter Wein*, or in the form of partitive and measure constructions, e.g., *ein Liter guter Wein*. All of these constructions have in common that they contain phrasal specifiers, which can be headed by a normal substantive category, i.e., an ordinary common or proper noun. As long as we do not provide an adequate mechanism according to which these substantive categories are turned into functional categories in our sense, these phrasal specifiers cannot take the role of a phrasal functional head of the entire nominal (or clausal projection).

We referred this problem to future research because we also view it a general HPSG problem. The closest we can get to a solution in current HPSG are lexical rules, which simply cannot be the ultimate wisdom. Whereas other frameworks appear to provide a broader range of standard options. For example, in categorial grammar the problem could be solved by type raising (which can be seen as a kind of lexical rule, though), Frameworks such as Tree Adjoining Grammar simply do not force us to make the assumption that the a local tree as a functor must be anchored in a lexical item which is equivalent to the head of the root of the tree.

However, in terms of our contribution to HPSG, we would have achieved preciously little, if it were only for posing new problems. The main results that we hope to have provided at this level are the reconstruction of the phrase structure schemata in terms of a system of parametrizable binary branching structures reflecting basic grammatical relations, as well as the theory of functional heads comprising among others the notions of functional completeness, as well as the relation between substantive and functional categories reflected in the split of head features. In both of these cases, our motivation for suggesting the changes was mainly rooted in our concern with German as a language with a structure which sometimes substantially differs from English. In spite of this foundation in a monolingual description, we are nevertheless confident that some of these suggested changes, such as for example the elimination of unary expansions in the case of the head-complement rules, could provide some interesting improvement of the current theory.[1]

Of course, we do not expect though that our all aspects of our approach will carry over equally well to all different kinds of languages. Thus, it may well be the case that English provides a rather hard ground for an application of the theory of functional heads, as it is appears to have a much higher number of constructions in which a functional category is missing altogether, such as for example reduced relative clauses or complementizer-less subordinate clauses. For a language, such as Japanese, the suggested changes of the theory appears to be almost ideal. In this language we find a high degree of freedom of the word order among the complements and adjuncts combined with a rather rigid head-final linearization, which is well supported by a consistently left-recursive binary branching analysis. In addition, Japanese employs a highly developed and systematic inventory of functional particles in the role of case- or clause-marker, which simply scream to be analyzed as functional heads. Quite similarly, Chinese shows some rather complex combinatorics of specifiers and classifiers which can be easily captured by a more elaborate system of minor features.

As to results which could be relevant on a general computational linguistics level, we attempted to present some methods for grammar engineering which are oriented along the requirements of compact and non-redundant encoding. Whether we have achieved our objectives in this area, is very difficult to judge. This is not so much because we are not fully convinced that we have presented a faultless and consistent analysis as far as the given fragment is concerned—such errors are fortunately easily detected in a computational environment, in particular if such a fragment is implemented and embedded into a larger grammar. The problem is rather the lack of commonly accepted general standards for what properties and restrictions a well-designed computational grammar should satisfy.

To pick out one illustrative example: shifting processes and descriptions from a structural level down to a lexical level is commonly regarded as a desirable property, which apparently stems from the time when the alternatives were structural transformations at runtime against lexical transformations at compile time. Nowadays, when hardly anybody even thinks about applying structural transformations, the stipulation of a lexical rule still has the reputation of a well accepted

[1] In fact, more recent and yet unpublished developments point towards a change of the "official" HPSG, in the direction of a recoding of rule schemata into binary branching structures.

and adequate method, such that we find a proliferation of lexical rules, which does not take into account that every ambiguity in the lexicon could eventually project into the structure and thus cause a computational overhead.

Of course, a split of structurally relevant information already in the lexicon can be extremely useful, if we have some robust and efficient means to pin down the correct lexical entry at the earliest point possible, say by means of a part-of-speech tagger. The most extreme situation, where such a lexicalization of information produces almost optimal results, is the case of super-tagging in Tree Adjoining Grammar. Since lexical entries in TAG are directly associated with a wealth of structural information, the determination of the correct lexical items by means of tagging will yield as a result all the necessary structural building blocks which only have to be put together by some simple mechanisms.[2] Unfortunately, HPSG does not have these properties. Thus, it is quite likely that in some frameworks the introduction of ambiguities in the lexicon with structurally repercussions at the moment is quite a bad idea, which appears to be definitely worse than the usage of underspecification and which may even be inferior to locating the information in a structural rule.

What makes it so difficult to decide at the moment what techniques and methods are the best, is that we do not have any reliable results yet. Grammar fragments of a noticeable size written in high level formalisms are only just about to emerge, and research in processing such fragments efficiently has been on the way for comparatively little time only. Thus, we simply do not know yet whether the best strategy to process a broad coverage HPSG as it is, for example by means of control mechanisms or explanation-based learning techniques, or whether the compilation and transformation into some lower or different formalism will yield the best results.

In the latter case, it may well be that certain generalizations (in the information theoretic sense) have to be spelt out and instantiated again for example in the form of a context-free backbone with thousands of rules and a correspondingly high number of lexical categories. Even in this case, we would nevertheless claim that the elimination of disjunctions in favour of underspecification, as well as other strategies that we suggested, contributes to good grammar engineering in terms of transparency and clarity of the grammar code.

Still, without question it would certainly be much much more helpful if the computational linguistic community could come to a consensus on such issues, and were it simply by deciding whether they are issues at all.

[2]Cf. [Joshi and Srinivas 1994]. The results presented in this paper have been substantially improved in the meantime by additional smoothing techniques. (p.c. Srinivas)

Bibliography

Abney, S. P. 1987. *The English Noun Phrase in its Sentential Aspects.* Cambridge, Mass.: Institute of Technology.

Baader, F., H.-J. Bürckert, B. Nebel, W. Nutt, and G. Smolka. 1993. On the Expressivity of Feature Logics with Negation, Functional Uncertainty, and Sort Equations. *Journal of Logic, Language and Information* 2:1–18.

Backofen, R., and H.-U. Krieger. 1993. The $\mathcal{TDL}/\mathcal{UD}i\mathcal{N}e$ system. In *Report of the EAGLES Workshop on Implemented Formalisms at DFKI, Saarbrücken*, ed. R. Backofen, H.-U. Krieger, S. P. Spackman, and H. Uszkoreit, 67–74. Saarbrücken.

Baltin, M. R. 1989. Heads and Projections. In *Alternative Conceptions of Phrase Structure*, ed. M. Baltin and A. Kroch, 1–16. Chicago: Chicago University Press.

Barton, G. E., R. C. Berwick, and E. S. Ristad. 1987. *Computational Complexity and Natural Language.* Cambridge, Mass.: MIT Press.

Bech, G. 1983. *Studien über das deutsche Verbum infinitum.* Tübingen: Max Niemeyer Verlag. 2nd edition.

Bloomfield, L. 1933. *Language.* London: George Allen and Unwin.

Borsley, R. 1989. Phrase-Structure Conception and the *Barriers* Conception of Clause Structure. *Linguistics* 27:843–863.

Borsley, R. D. 1987. Subjects and Complements in HPSG. CSLI-Report CSLI-87-107, CSLI, Stanford University, September.

Bouma, G. 1988. Modifiers and Specifiers in Categorial Unification Grammar. *Linguistics* 26:21–46.

Bouma, G., E. König, and H. Uszkoreit. 1988. A Flexible Graph-Unification Formalism and its Application to Natural Language Processing. *IBM Journal of Research and Development* 32(2):170–184.

Bresnan, J. 1978. A Realistic Transformational Grammar. In *Linguistic Theory and Psychological Reality*, ed. M. Halle, J. Bresnan, and A. G. Miller, chapter 1, 1–59. Cambridge, Mass.: MIT Press.

Bresnan, J. 1982a. Control and Complementation. In *The Mental Representation of Grammatical Relations*, ed. J. Bresnan, 282–390. Cambridge, Mass.: MIT Press.

Bresnan, J. (ed.). 1982b. *The Mental Representation of Grammatical Relations.* Cambridge, Mass.: MIT Press.

Bresnan, J. 1982c. The Passive in Lexical Theory. In *The Mental Representation of Grammatical Relations*, ed. J. Bresnan, 3–86. Cambridge, Mass.: MIT Press.

Calder, J., E. Klein, and H. Zeevat. 1988. Unification Categorial Grammar: A Concise, Extendable Grammar for Natural Language Processing. In *Proceedings of the 12th International Conference on Computational Linguistics COLING-88 and the 24th Annual Meeting of the Association for Computational Linguistics.* Budapest, Hungary, August.

Carpenter, B. 1992. *The Logic of Typed Feature Structures. With Applications to Unification Grammars, Logic Programs and Constraint Resolution.* Vol. 24 of Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.

Chomsky, N. 1972. *Syntactic Structures.* The Hague: Mouton.

Cooper, R. 1990. Specifiers, Complements and Adjuncts in HPSG. Research Paper EUCCS RP-42, Centre for Cognitive Science, Edinburgh.

Culy, C. 1985. The Complexity of the Vocabulary of Bambara. *Linguistics and Philosophy* 8(3):345–351.

Dörre, J., and M. Dorna. 1993. CUF – A Formalism for Linguistic Knowledge Representation. In *Computational Aspects of Constraint-Based Linguistic Descriptions*, 1–22. DYANA Deliverable R1.2.A.

Drach, E. 1939. *Grundgedanken der deutschen Satzlehre.* Frankfurt/Main.

Drosdowski, G. (ed.). 1984. *Grammatik der deutschen Gegenwartssprache.* Vol. 4 of DUDEN. Mannheim/Wien/Zürich: Bibliographisches Institut. 4th edition.

Eisele, A., and J. Dörre. 1986. A Lexical Functional Grammar System in Prolog. Stuttgart.

Emele, M. C., and R. Zajac. 1990. Typed Unification Grammars. In *Proceedings of the 13th International Conference on Computational Linguistics COLING-90.* Helsinki.

Engelkamp, J., G. Erbach, and H. Uszkoreit. 1992. Handling Linear Precedence Constraints by Unification. In *Proceedings of the 30th Meeting of the Association for Computational Linguistics*, 201–208. Newark, Del. ACL.

Erbach, G., and B. Krenn. 1994. Idioms and Support Verb Constructions. In *German in Head-Driven Phrase Structure Grammar*, ed. J. Nerbonne, K. Netter, and C. Pollard, Vol. 46 of Lecture Notes, 365–396. Stanford University: CSLI Publications.

Flickinger, D., and J. Nerbonne. 1991. Inheritance and Complementation: A Case Study of Easy Adjectives and Related Nouns. Research Report RR-91-30, DFKI, Saarbrücken, September.

Frank, A. 1994. Verb Second by Underspecification. In *KONVENS '94*, ed. H. Trost, 121–130. Berlin. Springer-Verlag.

Gazdar, G., E. Klein, G. Pullum, and I. Sag. 1985. *Generalized Phrase Structure Grammar.* Oxford: Basil Blackwell.

Haider, H. 1985. Der Rattenfängerei muß ein Ende gemacht werden. *Wiener Linguistische Gazette* 35-36:27–50.

Haider, H. 1987. Matching Projections. In *Constituent Structure*, ed. A. Cardinaletti, G. Cinque, and G. Giusti. GLOW Conference.

Haider, H. 1988. Die Struktur der Deutschen NP. *Zeitschrift für Sprachwissenschaft* 7(1):32–59.

Haider, H. 1993. *Deutsche Syntax Generativ. Vorstudien zur Theorie einer projektiven Grammatik.* Tübingen: Narr.

Haider, H., and K. Netter. 1991. Introduction. Derivation or Representation? In *Representation and Derivation in the Theory of Grammar. Studies in NLLT 22*, ed. H. Haider and K. Netter, 1–15. Dordrecht: Reidel.

Hepple, M. 1990. Verb Movement in German and Dutch. Technical report, Centre for Cognitive Science, Edinburgh.

Hinrichs, E., and T. Nakazawa. 1989. Flipped Out: Aux in German. In *Papers from the 25th Regional Meeting of the Chicago Linguistic Society*, Vol. 25, 193–202. Chicago, Illinois. CLS.

Hinrichs, E., and T. Nakazawa. 1994. Linearizing Finite AUX in German Complex VPs. In *German in Head-Driven Phrase Structure Grammar*, ed. J. Nerbonne, K. Netter, and C. Pollard, Vol. 46 of Lecture Notes, 11–37. Stanford University: CSLI Publications.

Hoeksema, J. 1985. Wazdat? - Contracted Forms and Verb Second in Dutch. In *Germanic Linguistics*, ed. J. Farlund, 112–124. Indiana. IULC.

Höhle, T. N. 1983. Topologische Felder. Köln.

Höhle, T. N. 1991. On Reconstruction and Coordination. In *Representation and Derivation in the Theory of Grammar. Studies in NLLT 22*, ed. H. Haider and K. Netter, 139–197. Dordrecht: Reidel.

Jackendoff, R. 1977. $\overline{\mathrm{X}}$ *Syntax: A Study of Phrase Structure.* Cambridge: MIT Press.

Johnson, M. 1986. The LFG Treatment of Discontinuity and the Double Infinitive Construction in Dutch. Research Report CSLI-86-65, CSLI, Stanford University.

Joshi, A. K., and B. Srinivas. 1994. Disambiguation of Super Parts of Speech (or Supertags): Almost Parsing. In *Proceedings of the 15th International Conference on Computational Linguistics, COLING-94*. Kyoto, Japan.

Kaplan, R., and J. Bresnan. 1982. Lexical Functional Grammar: A Formal System for Grammatical Representation. In *The Mental Representation of Grammatical Relations*, ed. J. Bresnan, chapter 4, 173–281. Cambridge, MA.: MIT Press.

Karttunen, L. 1984. Features and Values. In *Proceedings of the 10th International Conference on Computational Linguistics COLING-84*, 28–33.

Karttunen, L. 1989. Radical Lexicalism. In *Alternative Conceptions of Phrase Structure*, ed. M. Baltin and A. Kroch, 43–65. Chicago: Chicago University Press.

Kasper, R. 1994. Adjuncts in the Mittelfeld. In *German in Head-Driven Phrase Structure Grammar*, ed. J. Nerbonne, K. Netter, and C. Pollard, Vol. 46 of Lecture Notes, 39–69. Stanford University: CSLI Publications.

Kasper, R., B. Kiefer, K. Netter, and K. Vijay-Shanker. 1995. Compilation of HPSG to TAG. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 92–99. Cambridge, Mass. ACL.

Kathol, A. 1994. Passives without Lexical Rules. In *German in Head-Driven Phrase Structure Grammar*, ed. J. Nerbonne, K. Netter, and C. Pollard, Vol. 46 of Lecture Notes, 237–272. Stanford University: CSLI Publications.

Kay, M. 1979. Functional Grammar. In *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA. Berkeley Linguistics Society.

Kiss, T., and B. Wesche. 1991. Verb Order and Head Movement. In *Text Understanding in LILOG: Integrating Computational Linguistics and Artificial Intelligence. Final Report on the IBM Germany LILOG-Project*, ed. O. Herzog and C. Rollinger, 216–240. Berlin, Heidelberg: Springer.

Kolb, H. P., and C. Thiersch. 1991. Levels and Empty Categories in a Principles and Parameters Approach to Parsing. In *Representation and Derivation in the Theory of Grammar. Studies in NLLT 22*, ed. H. Haider and K. Netter, 251–299. Dordrecht: Reidel.

Kornai, A., and G. K. Pullum. 1990. The X-bar Theory of Phrase Structure. *Language* 66(1):24–50.

Krieger, H.-U. 1995. *TDL—A Type Description Language for Constraint-Based Grammars. Foundations, Implementation, and Applications.* PhD thesis, Universität des Saarlandes, Department of Computer Science, Saarbrücken.

Krieger, H.-U., and U. Schäfer. 1994a. $\mathcal{TDL}$—A Type Description Language for Constraint-Based Grammars. In *Proceedings of the 15th International Conference on Computational Linguistics, COLING-94*, 893–899. Kyoto, Japan.

Krieger, H.-U., and U. Schäfer. 1994b. $\mathcal{TDL}$—A Type Description Language for HPSG. Part 1: Overview. Research Report RR-94-37, DFKI, Saarbrücken.

Krieger, H.-U., and U. Schäfer. 1994c. $\mathcal{TDL}$—A Type Description Language for HPSG. Part 2: User Manual. DFKI Document D-94-14, DFKI, Saarbrücken.

Krieger, H.-U., and U. Schäfer. 1995. Efficient Parameterizable Type Expansion for Typed Feature Formalisms. In *Proceedings of the 14^{th} International Joint Conference on Artificial Intelligence (IJCAI).* Montreal, Canada.

LaPointe, S. 1990. Edge Features in GPSG. In *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, Vol. 26. Chicago, Illinois. CLS, CLS.

Maier, P., and P. Steffens. 1988. Determinatoren und Quantoren in einer kategorialen Unifikationsgrammatik des Deutschen. In *4. Österreichische Artificial-Intelligence-Tagung: Wiener Workshop - Wissensbasierte Sprachverarbeitung*, ed. H. Trost, 20–32. Berlin, Heidelberg: Springer.

Manandhar, S. 1995. Deterministic Consistency Checking of LP Constraints. In *Proceedings of the 7th European Conference of the Association for Computational Linguistics*, 165–172. Dublin. EACL.

Matthews, P. H. 1981. *Syntax.* Cambridge: Cambridge University Press.

Miller, P. H. 1992. *Clitics and Constituents in Phrase Structure Grammar.* New York: Garland. Published version of 1991 Doctoral dissertation, University of Utrecht, The Netherlands.

Nerbonne, J. 1986. 'Phantoms' and German Fronting: Poltergeist Constituents? *Linguistics* 24-25(285):857–870.

Nerbonne, J. 1994a. Edges and null nominal heads. In *HPSG-94 Conference*, Abstract. Kopenhagen.

Nerbonne, J. 1994b. Partial Verb Phrases and Spurious Ambiguities. In *German in Head-Driven Phrase Structure Grammar*, ed. J. Nerbonne, K. Netter, and C. Pollard, Vol. 46 of Lecture Notes, 109–150. Stanford University: CSLI Publications.

Netter, K. 1986. Getting Things Out of Order: An LFG-Proposal for the Treatment of German Word Order. In *Proceedings of the 11th International Conference on Computational Linguistics COLING-86*, 494–496. Bonn.

Netter, K. 1987. Wortstellung und Verbalkomplex in Deutschen. In *Computerlinguistik und Philologische Datenverarbeitung*, ed. U. Klenk, P. Scherber, and M. Thaller. Hildesheim: Olms.

Netter, K. 1991. Clause Union and Verb Raising Phenomena in German. Research Report RR-91-21, DFKI, Saarbrücken.

Netter, K. 1988. Non-Local Dependencies and Infinitival Constructions in German. In *Natural Language Parsing and Linguistic Theories*, ed. U. Reyle and C. Rohrer, 356–410. Dordrecht: Reidel.

Netter, K. 1992. On Non-Head Non-Movement. An HPSG Treatment of Finite Verb Position in German. In *Proceedings of KONVENS 92*, ed. G. Görz. Berlin/Heidelberg/New York: Springer.

Netter, K. 1993. Architecture and Coverage of the Disco Grammar. In *DFKI Workshop on Natural Language Systems: Re-usability and Modularity. Proceedings*, ed. S. Busemann and K. Harbusch, 1–10. Saarbrücken.

Netter, K., R. Kasper, B. Kiefer, and K. Vijay-Shanker. 1994. HPSG and TAG. In *In: 3e Colloque International sur les Grammaires d'Arbres Adjoints. (TAG+3). Rapport Technique TALANA-RT-94-01*, 77–82. Paris.

Netter, K., and J. Wedekind. 1986. An LFG-Based Approach to Machine Translation. In *Proceedings of IAI-MT-86*, 199–209. IAI, Saarbrücken.

Neumann, G. 1994. Application of Explanation-based Learning for Efficient Processing of Constraint-based Grammars. In *Proceedings of the Tenth IEEE Conference on Artifical Intelligence for Application.* San Antonio, Texas, March.

Oliva, K. 1992. Word Order Constraints in Binary Branching Syntactic Structures. CLAUS 20, Universität Saarbrücken, Computerlinguistik, Saarbrücken.

Pereira, F. C. N., and D. H. D. Warren. 1980. Definite Clause Grammars for Language Analysis—A Survey of the Formalism and a Comparison with Augmented Transition Networks. *Artificial Intelligence* 13:231–278.

Peters, S. P., and R. W. Ritchie. 1973. On the Generative Power of Transformational Grammars. *Information Sciences* 6:49–83.

Pollard, C. 1991. On Head Non–Movement. In *Proceedings of the Symposium on Discontinuous Constituency*, Appendix. Tilburg.

Pollard, C., and I. Sag. 1987. *Information-Based Syntax and Semantics. Vol. I: Fundamentals.* Vol. 13 of Lecture Notes. Stanford University: CSLI.

Pollard, C. J., and I. A. Sag. 1994. *Head Driven Phrase Structure Grammar.* Studies in Contemporary Linguistics. Chicago:University of Chicago Press & Stanford: CSLI Publications.

Pullum, G. K. 1991. English Nominal Gerund Phrases as Noun Phrases with Verb-Phrase Heads. *Linguistics* 29:763–799.

Reape, M. 1994. Domain Union and Word Order Variation in German. In *German in Head-Driven Phrase Structure Grammar*, ed. J. Nerbonne, K. Netter, and C. Pollard, Vol. 46 of Lecture Notes, 151–197. Stanford University: CSLI Publications.

Reis, M. 1980. On Justifying Topological Frames: 'Positional Field' and the Order of Nonverbal Constituents in German. *DRLAV* 22/23 (80):59–85.

Reis, M. 1985. Satzeinleitende Strukturen im Deutschen. Über COMP, Haupt- und Nebensätze, w-Bewegungen und die Doppelkopfanalyse. In *Erklärende Syntax des Deutschen*, ed. W. Abraham, Vol. 25 of Studien zur deutschen Grammatik, 271–311. Tübingen: Narr.

Sag, I. A. 1995. English Relative Clause Constructions. Unpublished ms., Stanford University.

Schäfer, U., and H.-U. Krieger. 1992. TDL *extra-light User's Guide: Franz Allegro Common LISP Version.* DISCO.

Scherpenisse, W. 1985. Die Satzstruktur des Deutschen und des Niederländischen im Rahmen der GB-Theorie. Eine Reaktion auf Marga Reis' Doppelkopfkritik. In *Erklärende Syntax des Deutschen*, ed. W. Abraham, Vol. 25 of Studien zur deutschen Grammatik, 313–334. Tübingen: Narr.

Sgall, P., E. Hajičová, and E. Benešová. 1973. *Topic, Focus and Generative Semantics.* Kronberg/Taunus: Scriptor.

Shieber, S., H. Uszkoreit, F. Pereira, J. Robinson, and M. Tyson. 1983. The Formalism and Implementation of PATR-II. In *Research on Interactive Acquisition and Use of Knowledge*, ed. B. J. Grosz and M. E. Stickel, 39–79. Menlo Park, Cal.: AI Center, SRI International.

Shieber, S. M. 1985. Evidence against the Context-Freeness of Natural Language. *Linguistics and Philosophy* 8:333–343.

Shieber, S. M. 1986. A Simple Reconstruction of GPSG. In *Proceedings of the 11th International Conference on Computational Linguistics COLING-86*, 211–215. Bonn.

Steedman, M. 1987. Combinatory Grammars and Parasitic Gaps. *Natural language and linguistic theory* 5:403–439.

Stowell, T. 1989. Subjects, Specifiers and X-Bar Theory. In *Alternative Conceptions of Phrase Structure*, ed. M. Baltin and A. Kroch, 232–262. Chicago: Chicago University Press.

Uszkoreit, H. 1986a. Categorical Unification Grammars. In *Proceedings of the 11th International Conference on Computational Linguistics COLING-86*, 187–194. Bonn.

Uszkoreit, H. 1986b. Linear Precedence in Discontinous Constituents. Complex Fronting in German. Report CSLI-86-47, Center for the Study of Language and Information, Stanford University, January.

Uszkoreit, H., R. Backofen, S. Busemann, A. K. Diagne, E. A. Hinkelman, W. Kasper, B. Kiefer, H.-U. Krieger, K. Netter, G. Neumann, S. Oepen, and S. P. Spackman. 1994. DISCO–An HPSG-based NLP System and its Application for Appointment Scheduling. In *Proceedings of the 15th International Conference on Computational Linguistics, COLING-94*. Kyoto, Japan.

van Riemsdijk, H. 1985. Zum Rattenfängereffekt bei Infinitiven in deutschen Relativsätzen. In *Erklärende Syntax des Deutschen*, ed. W. Abraham, Vol. 25 of Studien zur deutschen Grammatik, 75–98. Tübingen: Narr.

Weuster, E. 1983. Nicht-eingebettete Satztypen mit Verb-Endstellung im Deutschen. In *Zur Wortstellungsproblematik im Deutschen*, ed. K. Olszok and E. Weuster, 7–87. Tübingen: Narr.

Wunderlich, D. 1983. On the Compositionality of German Prefix Verbs. In *Meaning, Use and Interpretaion of Language*, ed. R. Bäuerle et al., 452–465. Berlin u. a.

Zajac, R. 1992. Inheritance and Constraint-Based Grammar Formalisms. *Computational Linguistics* 18(2):159–182.

Zwicky, A. M. 1986. German Adjective Agreement in GPSG. *Linguistics* 24(5):957–990.