# 4 Transformational Grammar – Minimalism

Like the Government & Binding framework that was introduced in the previous chapter, the Minimalist framework was initiated by Noam Chomsky at the MIT in Boston. Chomsky (1993, 1995b) argued that the problem of language evolution should be taken seriously and that the question of how linguistic knowledge could become part of our genetic endowment should be answered. To that end he suggested refocusing the theoretical developments towards models that have to make minimal assumptions regarding the machinery that is needed for linguistic analyses and hence towards models that assume less language specific innate knowledge.

Like GB, Minimalism is wide-spread: theoreticians all over the world are working in this framework, so the following list of researchers and institutions is necessarily incomplete. *Linguistic Inquiry* and *Syntax* are journals that almost exclusively publish Minimalist work and the reader is referred to these journals to get an idea about who is active in this framework. The most prominent researchers in Germany are Artemis Alexiadou, Humboldt University Berlin; Günther Grewendorf (2002), Frankfurt am Main; Joseph Bayer, Konstanz; and Gereon Müller, Leipzig.

While innovations like $\overline{\text{X}}$ theory and the analysis of clause structure in GB are highly influential and can be found in most of the other theories that are discussed in this book, this is less so for the technical work done in the Minimalist framework. It is nevertheless useful to familiarize with the technicalities since Minimalism is a framework in which a lot of work is done and understanding the basic machinery makes it possible to read empirically interesting work in that framework.

While the GB literature of the 1980s and 1990s shared a lot of assumptions, there was an explosion of various approaches in the Minimalist framework that is difficult to keep track of. The presentation that follows is based on David Adger's textbook (Adger 2003).

## 4.1 General remarks on the representational format

The theories that are developed in the framework of the Minimalist Program build on the work done in the GB framework. So a lot of things that were explained in the previous chapter can be taken over to this chapter. However, there have been some changes in fundamental assumptions. The general parameterized principles were dropped from the theory and instead the relevant distinctions live in features. Languages differ in the values that certain features may have and in addition to this, features may be strong or weak and feature strength is also a property that may vary from language to language.

Strong features make syntactic objects move to higher positions. The reader is familiar with this feature-driven movement already since it was a component of the movement-based analysis of the passive in Section 3.4. In the GB analysis of passive, the object had to move to the specifier position of IP in order to receive case. Such movements that are due to missing feature values are a key component in Minimalist proposals.

### 4.1.1 Basic architecture

Chomsky assumes that there are just two operations (rules) for combining linguistic objects: External and Internal Merge. External Merge simply combines two elements like *the* and *book* and results in a complex phrase. Internal Merge is used to account for movement of constituents. It applies to one linguistic object and takes some part of this linguistic object and adjoins it to the left of the respective object. The application of External Merge and Internal Merge can apply in any order. For instance, two objects can be combined with External Merge and then one of the combined items is moved to the left by applying Internal Merge. The resulting object can be externally merged with another object and so on. As an example consider the NP in (1):

(1) the man who we know

To derive this NP the verb *know* is externally merged with its object *who*. After several intermediate merges that will be discussed below, *know who* will be merged with *we* and finally the *who* is moved to the left by Internal Merge, resulting in *who we know*. This relative clause can be externally merged with *man* and so on.

So, Minimalist theories differ from GB in not assuming a Deep Structure that is generated by some $\overline{\text{X}}$ grammar and a Surface Structure that is derived from the Deep Structure by move $\alpha$. Instead, it is assumed that there is a phase in which External and Internal Merge (combination and movement) apply in any order to derive a certain structure that is then said to be spelled out. It is said that the structure is sent to the interfaces: the articulatory-perceptual system (AP) on the one hand and the conceptual-intentional system (CI) on the other side. AP corresponds to the level of Phonological Form (PF) and CI to the level of Logical Form (LF) in GB. The new architecture is depicted in Figure 4.1 on the next page. Overt syntax stands for syntactic operations that usually have a visible effect. After overt syntax the syntactic object is sent off to the interfaces and some transformations may take place after this Spell-Out point. Since such transformations do not affect pronunciation, this part of syntax is called *covert syntax*. Like in GB's LF, the covert syntax can be used to derive certain scope readings.

This architecture was later modified to allow Spell-Out at several points in the derivation. It is now assumed that there are *phases* in a derivation and that a completed phase is spelled out once it is used in a combination with a head (Chomsky 2008). For instance, a subordinated sentence like *that Peter comes* in (2) is one phase and is sent to the interfaces before the whole sentence is completed.[1]

[1] Andreas Pankau (p. c. 2015) pointed out to me that there is a fundamental problem with such a conception of phases, since if it is the case that only elements that are in a relation to a head are send off to the interface

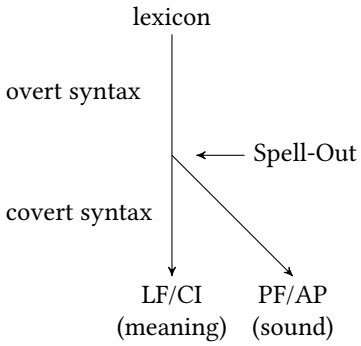



Figure 4.1: Architecture assumed in Minimalist theories before the Phase model

(2) He believes that Peter comes.

There are different proposals as to what categories form complete phases. Since the concept of phases is not important for the following introduction, I will ignore this concept in the following. See Section 15.1 on the psycholinguistic plausibility of phases in particular and the Minimalist architecture in general.

### 4.1.2 Valence, feature checking, and agreement

The basic mechanism in Minimalist theories is feature checking. For instance, the noun *letters* may have a P feature, which means that it has to combine with a PP in order to form a complete phrase.

(3) letters to Peter

It is assumed that there are interpretable and uninterpretable features. An example of an interpretable feature is the number feature of nouns. The singular/plural distinction is semantically relevant. The category features for part of speech information are purely syntactic and hence cannot be interpreted semantically. Minimalism assumes that all uninterpretable features have to be used up during the derivation of a complex linguistic object. This process of eating up the features is called *checking*. As an example, let us consider the noun *letters* again. The analysis of (3) is depicted in Figure 4.2 on the following page. The fact that the P feature of *letters* is uninterpretable is represented by the little *u* in front of the P. The uninterpretable P feature of *letters* can be checked against the P feature of *to Peter*. All checked features are said to delete automatically. The deletion is marked by striking the features out in the figures. Strings like (4) are ruled out as complete derivations since the N feature of P is not checked. This situation is shown in Figure 4.3 on the next page.

then the top-most phrase in a derivation would never be sent to the interfaces, since it does not depend on any head.

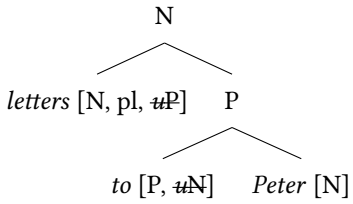



Figure 4.2: Valence representation via uninterpretable features

(4) * letters to

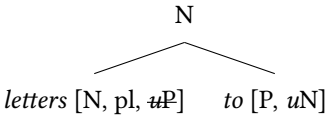



Figure 4.3: Illegitimate syntactic object due to an uninterpretable feature

If this structure would be used in a larger structure that is spelled out, the derivation would *crash* since the conceptual system could not make sense of the N feature that is still present at the P node.

Selectional features are atomic, that is, the preposition cannot select an NP[*acc*] as in GB and the other theories in this book unless NP[*acc*] is assumed to be atomic. Therefore, an additional mechanism is assumed that can check other features in addition to selectional features. This mechanism is called *Agree*.

(5) a. * letters to he
b. letters to him

The analysis of (5b) is shown in Figure 4.4. There is an interesting difference between the

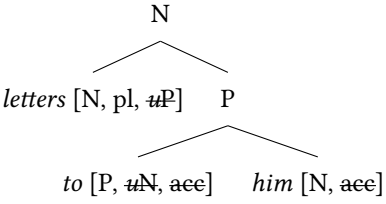



Figure 4.4: Feature checking via Agree

checking of selectional features and the checking of features via Agree. The features that

are checked via Agree do not have to be at the top node of the object that is combined with a head. This will play a role later in the analysis of the passive and local reordering.

### 4.1.3 Phrase structure and $\overline{\text{X}}$ theory

The projections of $\overline{\text{X}}$ structures were given in Figure 2.9 on page 78. According to early versions of the $\overline{\text{X}}$ theory, there could be arbitrarily many complements that were combined with $\text{X}^0$ to form an $\overline{\text{X}}$. Arbitrarily many adjuncts could attach to $\overline{\text{X}}$ and then at most one specifier could be combined with the $\overline{\text{X}}$ yielding an XP. Minimalist theories assume binary branching and hence there is at most one complement, which is the first-merged item. Furthermore, it is not assumed that there is a unique specifier position. Chomsky rather assumes that all items that are not complements are specifiers. That is he distinguishes between first-merged (complements) and later-merged items (specifiers). Figure 4.5 shows an example with two specifiers. It is also possible to have just

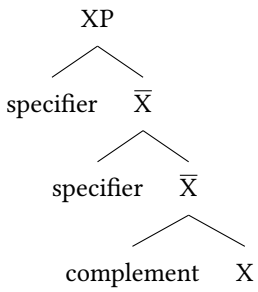



Figure 4.5: Complements and specifiers in Minimalist theories

a complement and no specifier or to have one or three specifiers. What structures are ultimately licensed depends on the features of the items that are involved in the Merge operations. Whether a phrasal projection counts as an $\overline{\text{X}}$ or an XP depends on whether the phrase is used as a complement or specifier of another head or whether it is used as head in further Merge operations. If a phrase is used as specifier or complement its status is fixed to be a phrase (XP), otherwise the projectional status of resulting phrases is left underspecified. Lexical head daughters in Merge operations have the category X and complex head daughters in Merge operations have the category $\overline{\text{X}}$. This solves the problem that standard $\overline{\text{X}}$ theoretic approaches had with pronouns and proper names: a lot of unary branching structure had to be assumed (See left picture in Figure 2.9). This is not necessary any longer in current Minimalist theories.[2]

[2] For problems with this approach see Brosziewski (2003: Chapter 2.1).

### 4.1.4 Little *v*

In Section 3.4 I used $\overline{\text{X}}$ structures in which a ditransitive verb was combined with its accusative object to form a $\overline{\text{V}}$, which was then combined with the dative object to form a further $\overline{\text{V}}$. Such binary branching structures and also flat structures in which both objects are combined with the verb to form a $\overline{\text{V}}$ are rejected by many practitioners of GB and Minimalism since the branching does not correspond to branchings that would be desired for phenomena like the binding of reflexives and negative polarity items. A binding in which *Benjamin* binds *himself* in (6a) is impossible:

(6) a. * Emily showed himself Benjamin in the mirror.
b. Peter showed himself Benjamin in the mirror.

What is required for the analysis of Binding and NPI phenomena in theories that analyze these phenomena in terms of tree configurations is that the reflexive pronoun is "higher" in the tree than the proper name *Benjamin*. More precisely, the reflexive pronoun *himself* has to c-command *Benjamin*. c-command is defined as follows (Adger 2003: 117):[3]

(7) A node A c-commands B if, and only if A's sister either:
a. is B, or
b. contains B

In the trees to the left and in the middle of Figure 4.6 the c-command relations are not as desired: in the left-most tree both NPs c-command each other and in the middle one *Benjamin* c-commands *himself* rather than the other way round. Hence it is assumed

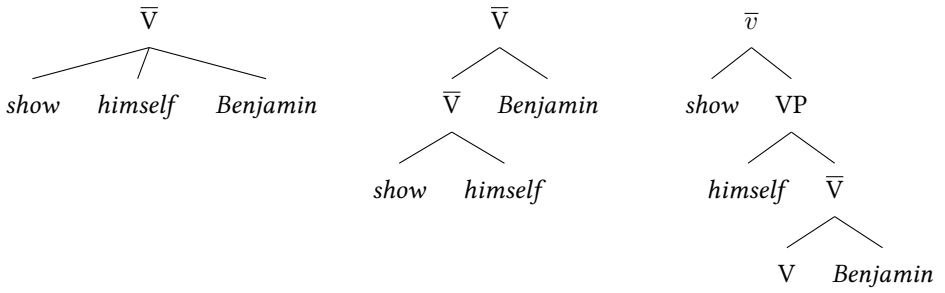



Figure 4.6: Three possible analyses of ditransitives

that the structures at the left and in the middle are inappropriate and that there is some additional structure involving the category *v*, which is called *little v* (Adger 2003: Section 4.4). The sister of *himself* is $\overline{\text{V}}$ and $\overline{\text{V}}$ contains *Benjamin*, hence *himself* c-commands

[3] c-command also plays a prominent role in GB. In fact, one part of Government & Binding is the Binding Theory, which was not discussed in the previous chapter since binding phenomena do not play a role in this book.

*Benjamin*. Since the sister of *Benjamin* is V and V neither is nor contains *himself*, *Benjamin* does not c-command *himself*.

The analysis of ditransitives involving an additional verbal head goes back to Larson (1988). Hale & Keyser (1993a: 70) assume that this verbal head contributes a causative semantics. The structure in Figure 4.7 is derived by assuming that the verb *show* starts out in the V position and then moves to the *v* position. *show* is assumed to mean *see* and in the position of little *v* it picks up the causative meaning, which results in a *cause-see'* meaning (Adger 2003: 133).

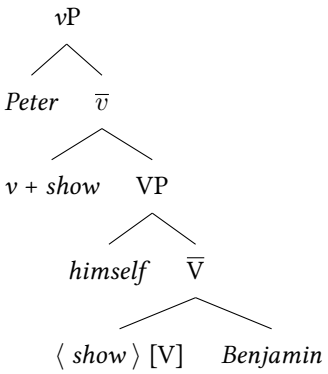



Figure 4.7: Analysis of ditransitives involving movement to little *v*

While the verb shell analysis with an empty verbal head was originally invented by Larson (1988) for the analysis of ditransitive verbs, it is now also used for the analysis of strictly transitive and even intransitive verbs.

Adger (2003: Section 4.5) argues that semantic roles are assigned uniformly in certain tree configurations:

(8) a. NP daughter of *v*P → interpreted as agent
b. NP daughter of VP → interpreted as theme
c. PP daughter of $\overline{v}$ → interpreted as goal

Adger assumes that such uniformly assigned semantic roles help in the process of language acquisition and from this it follows that little *v* should also play a role in the analysis of examples with strictly transitive and intransitive verbs. The Figures 4.8 and 4.9 show the analysis of sentences containing the verbs *burn* and *laugh* respectively.[4]

Adger (2003: 164) assumes that intransitive and transitive verbs move from V to little *v* as well. This will be reflected in the following figures.

[4] If all intransitive verbs of this type are supposed to have agents as subjects, a very broad conception of agent has to be assumed that also subsumes the subject of verbs like *sleep*. Usually sleeping is not an activity that is performed intentionally.

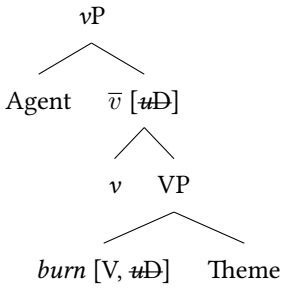



Figure 4.8: Analysis of strictly transitives involving little *v*

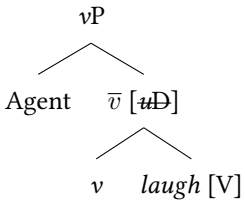



Figure 4.9: Analysis of intransitives involving little *v*

### 4.1.5 CP, TP, *v*P, VP

Section 3.1.5 dealt with the CP/IP system in GB. In the course of the development of Minimalism, the Inflectional Phrase was split into several functional projections (Chomsky 1989) of which only the Tense Phrase is assumed in current Minimalist analyses. So, the TP of Minimalism corresponds to IP in the GB analysis. Apart from this change, the core ideas of the CP/IP analysis have been transferred to the Minimalist analysis of English. This subsection will first discuss special features that are assumed to trigger movement (Subsection 4.1.5.1) and then case assignment (Subsection 4.1.5.2).

#### 4.1.5.1 Features as triggers for movement: The EPP feature on T

In GB approaches, the modals and auxiliaries were analyzed as members of the category I and the subjects as specifiers of IP. In the previous section, I showed how subjects are analyzed as specifiers of *v*P. Now, if one assumes that a modal verb combines with such a *v*P, the subject follows the modal, which does not correspond to the order that is observable in English. This problem is solved by assuming a strong uninterpretable D feature at T. Since the feature is strong, a suitable D has to move to the specifier of T and check the D locally. Figure 4.10 on the next page shows the TP that plays a role in the analysis of (9):

(9) Anna will read the book.

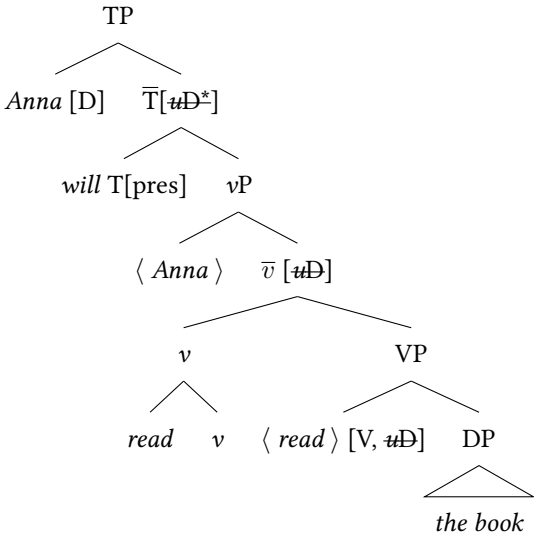



Figure 4.10: Analysis of *Anna will read the book.* involving a modal and movement of the subject from *v* to T

The DP *the book* is the object of *read* and checks the D feature of *read*. little *v* selects for the subject *Anna*. Since T has a strong D feature (marked by an asterisk '*'), *Anna* must not remain inside of the *v*P but moves on to the specifier position of TP.

Full sentences are CPs. For the analysis of (9), an empty C head is assumed that is combined with the TP. The empty C contributes a clause type feature Decl. The full analysis of (9) is shown in Figure 4.11.

The analysis of the question in (10) involves an unvalued clause-type feature on T for the sentence type *question*.

(10) What will Anna read?

The empty complementizer C has a Q feature that can value the clause-type feature on T. Since clause-type features on T that have the value Q are stipulated to be strong, the T element has to move to C to check the feature locally. In addition, the *wh* element is moved. This movement is enforced by a strong wh feature on C. The analysis of (10) is given in Figure 4.12 on page 139.

#### 4.1.5.2 Case assignment

In the GB analysis that was presented in Chapter 3, nominative was assigned by (finite) I and the other cases by the verb (see Section 3.4.2). The assignment of nominative is taken over to Minimalist analyses, so it is assumed that nominative is assigned by (finite) T. However, in the Minimalist theory under consideration, there is not a single verb

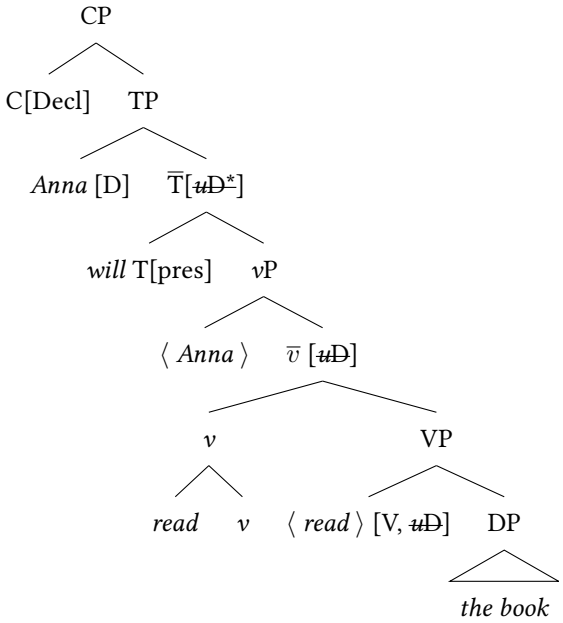



Figure 4.11: Analysis of *Anna will read the book.* as CP with an empty C with the clause-type feature Decl

projection, but there are two verbal projections: *v*P and VP. Now, one could assume that V assigns accusative to its complement or that *v* assigns accusative to the complement of the verb it dominates. Adger (2003: Section 6.3.2, Section 6.4) assumes the latter approach, since it is compatible with the analysis of so-called unaccusative verbs and the passive. Figure 4.13 on the next page shows the TP for (11):

(11) Anna reads the book.

The two NPs *Anna* and *the book* start out with unvalued uninterpretable case features: [*u*case:]. The features get valued by T and *v*. It is assumed that only one feature is checked by Merge, so this would be the D feature on T, leaving the case feature for the other available checking mechanism: Agree. Agree can be used to check features in sister nodes, but also features further away in the tree. The places that are possible candidates for Agree relations have to stand in a certain relation to each other. The first node has to c-command the node it Agrees with. c-command roughly means: one node up and then arbitrarily many nodes down. So *v* c-commands VP, V, the DP *the book*, and all the nodes within this DP. Since Agree can value features of c-commanded nodes, the accusative on *v* can value the case feature of the DP *the book*.

The non-locality that is build into Agree raises a problem: why is it that (12) is ungrammatical?

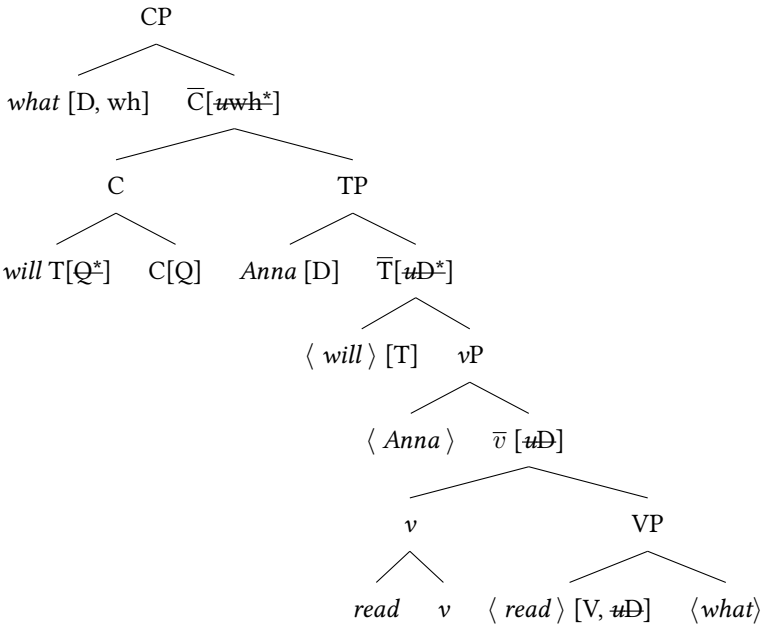



Figure 4.12: Analysis of *What will Anna read?* with an empty C with a strong wh feature

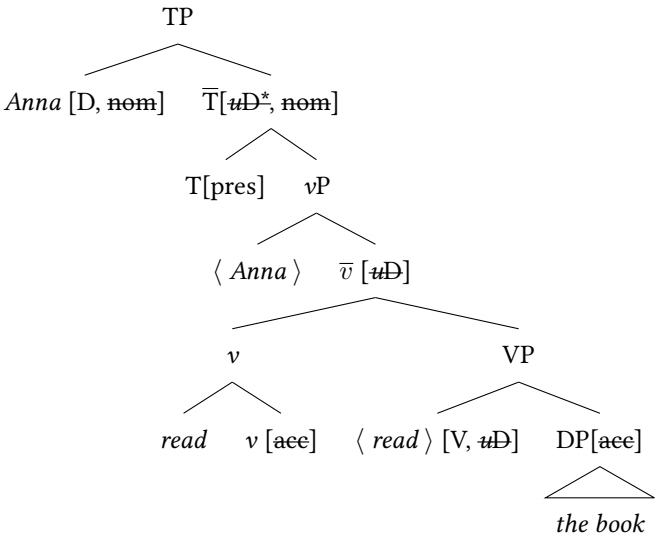



Figure 4.13: Case assignment by T and *v* in the TP for of *Anna reads the book.*

(12) * Him likes she.

The accusative of *v* could be checked with its subject and the nominative of T with the object of *likes*. Both DPs stand in the necessary c-command relations to T and *v*. This problem is solved by requiring that all Agree relations have to involve the closest possible element. Adger (2003: 218) formulates this constraint as follows:

(13) Locality of matching: Agree holds between a feature F on X and a matching feature F on Y if and only if there is no intervening Z[F].

Intervention is defined as follows:

(14) Intervention: In a structure [X … Z … Y], Z intervenes between X and Y iff X c-commands Y.

So, since T may Agree with *Anna* it must not Agree with *the book*. Hence nominative assignment to *she* in (12) is impossible and (12) is correctly ruled out.

### 4.1.6 Adjuncts

Adger (2003: Section 4.2.3) assumes that adjuncts attach to XP and form a new XP. He calls this operation *Adjoin*. Since this operation does not consume any features it is different from External Merge and hence a new operation would be introduced into the theory, contradicting Chomsky's claim that human languages use only Merge as a structure building operation. There are proposals to treat adjuncts as elements in special adverbial phrases with empty heads (see Section 4.6.1) that are also assumed to be part of a hierarchy of functional projections. Personally, I prefer Adger's solution that corresponds to what is done in many other frameworks: there is a special rule or operation for the combination of adjuncts and heads (see for instance Section 9.1.7 on the HPSG schema for head adjunct combinations).

## 4.2 Verb position

The analysis of verb first sentences in German is straightforward, given the machinery that was introduced in the previous section. The basic idea is the same as in GB: the finite verb moves from V to *v* to T and then to C. The movement to T is forced by a strong tense feature on T and the movement of the T complex to C is enforced by a clause-type feature on T that is valued as a strong Decl by C. The analysis of (15) is shown in Figure 4.14 on the facing page.

(15) Kennt jeder diesen Mann?
knows everybody this man
'Does everybody know this man?'

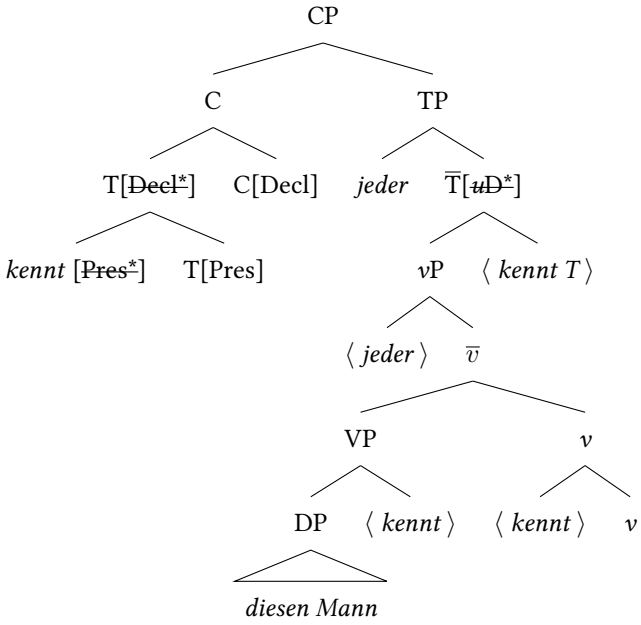



Figure 4.14: Analysis of *Kennt jeder diesen Mann?* 'Does everybody know this man?' following the analysis of Adger (2003)

## 4.3 Long-distance dependencies

Having explained the placement of the verb in initial position, the analysis of V2 sentences does not come with a surprise: Adger (2003: 331) assumes a feature that triggers the movement of a constituent to a specifier position of C. Adger calls this feature top, but this is a misnomer since the initial position in German declarative sentences is not restricted to topics. Figure 4.15 on the next page shows the analysis of (16):

(16) Diesen Mann kennt jeder.
this man knows everybody
'Everbody knows this man.'

## 4.4 Passive

Adger (2003) suggests an analysis for the passive in English, which I adapted here to German. Like in the GB analysis that was discussed in Section 3.4 it is assumed that the verb does not assign accusative to the object of *schlagen* 'to beat'. In Minimalist terms, this means that little *v* does not have an acc feature that has to be checked. This

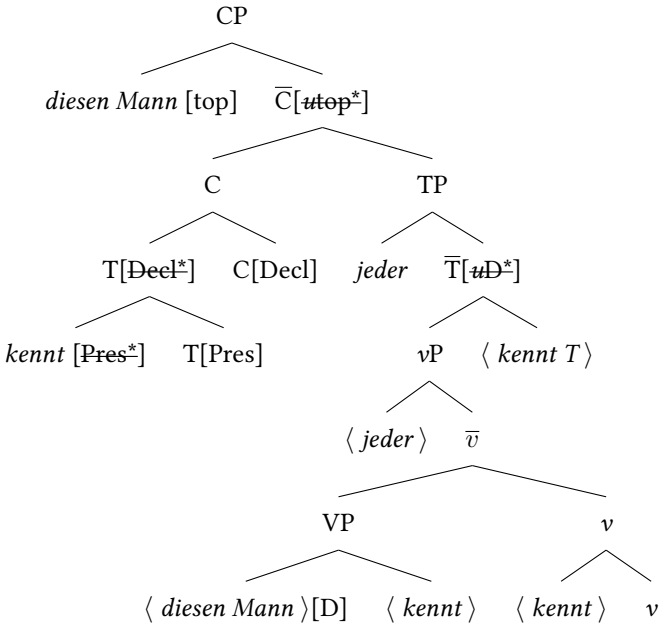



Figure 4.15: Analysis of *Diesen Mann kennt jeder.* 'This man, everybody knows.' following the analysis of Adger (2003: 331)

special version of little *v* is assumed to play a role in the analysis of sentences of so-called unaccusative verbs (Perlmutter 1978). Unaccusative verbs are a subclass of intransitive verbs that have many interesting properties. For instance, they can be used as adjectival participles although this is usually not possible with intransitive verbs:

(17) a. * der getanzte Mann
the danced man

b. der gestorbene Mann
the died man
'the dead man'

The explanation of this difference is that adjectival participles predicate over what is the object in active sentences:

(18) a. dass der Mann das Buch gelesen hat
that the man the book read has
'that the man read the book'

b. das gelesene Buch
the read book

Now the assumption is that the argument of *gestorben* 'died' behaves like an object, while the argument of *getanzt* 'danced' behaves like a subject. If adjectival passives predicate over the object it is explained why (17b) is possible, while (17a) is not.

Adger (2003: 140) assumes the structure in Figure 4.16 for *v*Ps with unaccusative verbs. It is assumed that this unaccusative variant of little *v* plays a role in the analysis of the

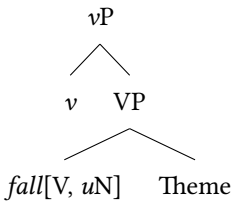



Figure 4.16: Structure of *v*P with unaccusative verbs like *fall*, *collapse*, *wilt* according to Adger (2003: 140)

passive. Unaccusative verbs are similar to passivized verbs in that they do have a subject that somehow also has object properties. The special version of little *v* is selected by the Passive head *werden* 'be', which forms a Passive Phrase (abbreviated as PassP). See Figure 4.17 for the analysis of (19):

(19) dass er geschlagen wurde
that he beaten was
'that he was beaten'

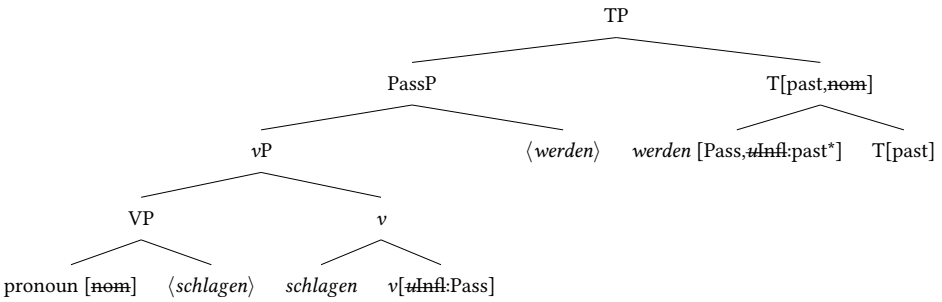



Figure 4.17: Minimalist analysis of the passive without movement but with nonlocal case assignment via Agree

The Pass head requires the Infl feature of little *v* to have the value Pass, which results in participle morphology at spellout. Hence the form that is used is *geschlagen* 'beaten'. The auxiliary moves to T to check the strong Infl feature at T and since the Infl feature is past, the past form of *werden* 'be', namely *wurde* 'was', is used at spellout. T has

a nom feature that has to be checked. Interestingly, the Minimalist approach does not require the object of *schlagen* to move to the specifier position of T in order to assign case, since case assignment is done via Agree. Hence the pronominal argument of *schlagen* in principle could stay in its object position and nevertheless get nominative from T. This would solve the problem of the GB analysis that was pointed out by Lenerz (1977: Section 4.4.3). See page 113 for Lenerz' examples and discussion of the problem. However, Adger (2003: 332) assumes that German has a strong EPP feature on T. If this assumption is upheld, all problems of the GB account will carry over to the Minimalist analysis: all objects have to move to T even when there is no reordering taking place. Furthermore, impersonal passives of the kind in (20) would be problematic, since there is no noun phrase that could be moved to T in order to check the EPP feature:

(20) weil getanzt wurde
because danced was
'because there was dancing there'

## 4.5 Local reordering

Adger (2003) does not treat local reordering. But there are several other suggestions in the literature. Since all reorderings in Minimalist theories are feature-driven, there must be an item that has a feature that triggers reorderings like those in (21b):

(21) a. [weil] jeder diesen Mann kennt
because everyone this man knows
'because everyone knows this man'
b. [weil] diesen Mann jeder kennt
because this man everyone knows

There have been various suggestions involving functional projections like Topic Phrase (Laenzlinger 2004: 222) or AgrS and AgrO (Meinunger 2000: Chapter 4) that offer places to move to. G. Müller (2014a: Section 3.5) offers a leaner solution, though. In his approach, the object simply moves to a second specifier position of little *v*. The analysis is depicted in Figure 4.18 on the facing page.[5]

An option that was suggested by Laenzlinger (2004: 229–230) is to assume several Object Phrases for objects that may appear in any order. The objects move to the specifier positions of these projections and since the order of the Object Phrases is not restricted, both orders in (22) can be analyzed:

(22) a. dass Hans diesen Brief meinem Onkel gibt
that Hans this letter my uncle gives
'that Hans gives this letter to my uncle'

[5] G. Müller assumes optional features on *v* and V that trigger local reorderings (p. 48). These are not given in the figure.

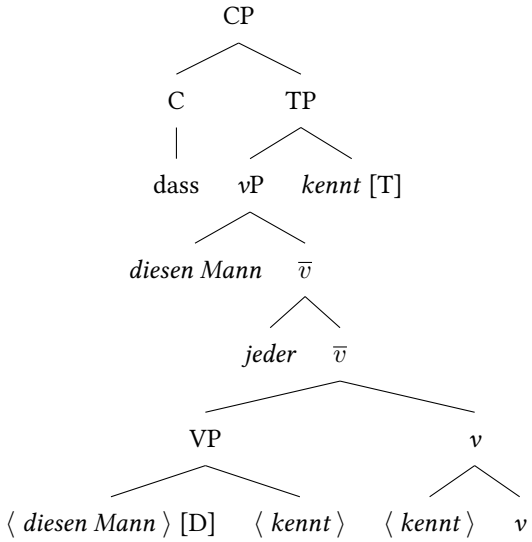



Figure 4.18: Analysis of *dass diesen Mann jeder kennt* 'that everybody knows this man' as movement of the object to a specifier position of *v*

b. dass Hans meinem Onkel diesen Brief gibt
that Hans my uncle this letter gives
'that Hans gives to my uncle this letter'

## 4.6 New developments and theoretical variants

At the start of the 90s, Chomsky suggested a major rethink of the basic theoretical assumptions of GB and only keeping those parts of the theory which are absolutely necessary. In the *Minimalist Program*, Chomsky gives the central motivations for the far-reaching revisions of GB theory (Chomsky 1993, 1995b). Until the beginning of the 90s, it was assumed that Case Theory, the Theta Criterion, $\overline{\text{X}}$ theory, Subjacency, Binding Theory, Control Theory etc. all belonged to the innate faculty for language (Richards 2015a: 804). This, of courses, begs the question of how this very specific linguistic knowledge made its way into our genome. The Minimalist Program follows up on this point and attempts to explain properties of language through more general cognitive principles and to reduce the amount of innate language-specific knowledge postulated. The distinction between Deep Structure and Surface Structure, for example, was abandoned. Move still exists as an operation, but can be used directly to build sub-structures rather than after a complete D-structure has been created. Languages differ with regard to whether this movement is visible or not.

Although Chomsky's Minimalist Program should be viewed as a successor to GB, advocates of Minimalism often emphasize the fact that Minimalism is not a theory as such, but rather a research program (Chomsky: 2007: 4; 2013: 6). The actual analyses suggested by Chomsky (1995b) when introducing the research program have been reviewed by theoreticians and have sometimes come in for serious criticism (Kolb 1997; Johnson & Lappin 1997, 1999; Lappin, Levine & Johnson 2000a,b, 2001; Seuren 2004; Pinker & Jackendoff 2005), however, one should say that some criticisms overshoot the mark.

There are various strains of Minimalism. In the following sections, I will discuss some of the central ideas and explain which aspects are regarded problematic.

### 4.6.1 Move, Merge, feature-driven movement and functional projections

Johnson, Lappin and Kolb have criticized the computational aspects of Chomsky's system. Chomsky suggested incorporating principles of economy into the theory. In certain cases, the grammatical system can create an arbitrary number of structures, but only the most economical, that is, the one which requires the least effort to produce, will be accepted as grammatical (transderivational economy). This assumption does not necessarily have to be taken too seriously and, in reality, does not play a role in many works in the Minimalist framework (although see Richards (2015b) for recent approaches with derivations which are compared in terms of economy). Nevertheless, there are other aspects of Chomsky's theory which can be found in many recent works. For example, Chomsky has proposed reducing the number of basic, structure building operations which license structures to two: Move and Merge (that is, Internal and External Merge). Move corresponds to the operation move $\alpha$, which was already discussed in Chapter 3, and Merge is the combination of (two) linguistic objects.

It is generally assumed that exactly two objects can be combined (Chomsky 1995b: 226). For Move, it is assumed that there must be a reason for a given movement operation. The reason for movement is assumed to be that an element can check some feature in the position it is moved to. This idea was already presented in the analysis of the passive in Section 3.4: the accusative object does not bear case in passive sentences and therefore has to be moved to a position where it can receive case. This kind of approach is also used in newer analyses for a range of other phenomena. For example, it is assumed that there are phrases whose heads have the categories focus and topic. The corresponding functional heads are always empty in languages like German and English. Nevertheless, the assumption of these heads is motivated by the fact that other languages possess markers which signal the topic or focus of a sentence morphologically. This argumentation is only possible if one also assumes that the inventory of categories is the same for all languages. Then, the existence of a category in one language would suggest the existence of the same category in all other languages. This assumption of a shared universal component (Universal Grammar, UG) with detailed language-specific knowledge is, however, controversial and is shared by few linguists outside of the Chomskyan tradition. Even for those working in Chomskyan linguistics, there have been questions

raised about whether it is permissible to argue in this way since if it is only the ability to create recursive structures that is responsible for the human-specific ability to use language (faculty of language in the narrow sense) – as Hauser, Chomsky & Fitch (2002) assume –, then the individual syntactic categories are not part of UG and data from other languages cannot be used to motivate the assumption of invisible categories in another language.

#### 4.6.1.1 Functional projections and modularization of linguistic knowledge

The assumption that movement must be licensed by feature checking has led to an inflation of the number of (silent) functional heads.[6] Rizzi (1997: 297) suggests the structure in Figure 4.19 on the next page (see Grewendorf 2002: 85, 240; 2009, too). The functional categories Force, Top, Foc and Fin correspond to clause type, topic, focus and finiteness. It is assumed that movement always targets a specifier position. Topics and focused elements are always moved to the specifier position of the corresponding phrase. Topics can precede or follow focused elements, which is why there are two topic projections: one above and one below FocP. Topic phrases are recursive, that is, an arbitrary number of TopPs can appear at the positions of TopP in the figure. Following Grewendorf (2002: 70), topic and focus phrases are only realized if they are required for particular information structural reasons such as movement.[7] Chomsky (1995b: 147) follows Pollock (1989) in assuming that all languages have functional projections for subject and object agreement as well as negation (AgrS, AgrO, Neg).[8] Sternefeld (1995: 78), von Stechow (1996: 103) and Meinunger (2000: 100–101, 124) differentiate between two agreement positions for direct and indirect objects (AgrO, AgrIO). As well as AgrS, AgrO and Neg, Beghelli & Stowell (1997) assume the functional heads Share and Dist in order to explain scope phenomena in English as feature-driven movements at LF. For a treatment of scope phenomena without empty elements or movement, see Section 19.3. Błaszczak & Gärtner (2005: 13) assume the categories −PolP, +PolP and %PolP for their discussion of polarity.

Webelhuth (1995: 76) gives an overview of the functional projections that had been proposed up to 1995 and offers references for AgrA, AgrN, AgrV, Aux, Clitic Voices, Gender, Honorific, $\mu$, Number, Person, Predicate, Tense, Z.

In addition to AdvP, NegP, AgrP, FinP, TopP and ForceP, Wiklund, Hrafnbjargarson, Bentzen & Hróarsdóttir (2007) postulate an OuterTopP.

[6] The assumption of such heads is not necessary since features can be 'bundled' and then they can be checked together. For an approach in this vein, which is in essence similar to what theories such as HPSG assume, see Sternefeld (2006: Section II.3.3.4, Section II.4.2).

In so-called cartographic approaches, it is assumed that every morphosyntactic feature corresponds to an independent syntactic head (Cinque & Rizzi 2010: 54, 61). For an explicitly formalized proposal in which exactly one feature is consumed during a combination operation see Stabler (2001: 335). Stabler's *Minimalist Grammars* are discussed in more detail in Section 4.6.4.

[7] There are differing opinions as to whether functional projections are optional or not. Some authors assume that the complete hierarchy of functional projections is always present but functional heads can remain empty (e. g. Cinque 1999: 106 and Cinque & Rizzi 2010: 55).

[8] See Chomsky (1995b: Section 4.10.1), however.

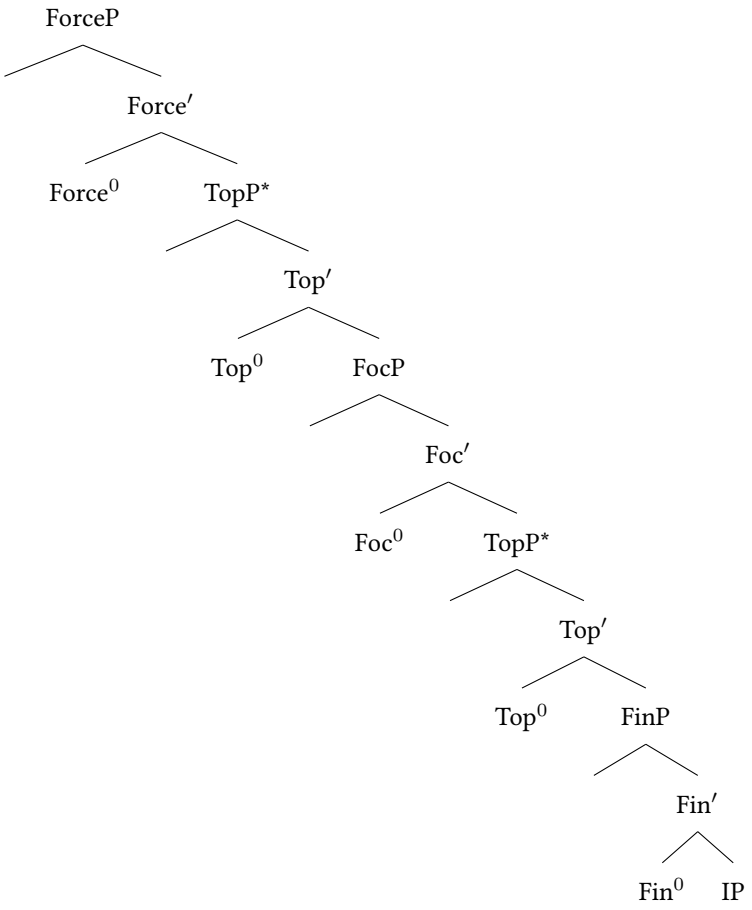



Figure 4.19: Syntactic structure of sentences following Rizzi (1997: 297)

Poletto (2000: 31) suggests both a HearerP and a SpeakerP for the position of clitics in Italian.

Cinque (1999: 106) adopts the 32 functional heads in Table 4.1 in his work. He assumes

Table 4.1: Functional heads following Cinque (1999: 106)

| | | | |
|---|---|---|---|
| 1. $\text{Mood}_{\text{Speech Act}}$ | 2. $\text{Mood}_{\text{Evaluative}}$ | 3. $\text{Mood}_{\text{Evidential}}$ | 4. $\text{Mood}_{\text{Epistemic}}$ |
| 5. T(Past) | 6. T(Future) | 7. $\text{Mood}_{\text{Irrealis}}$ | 8. $\text{Mod}_{\text{Necessity}}$ |
| 9. $\text{Mod}_{\text{Possibility}}$ | 10. $\text{Mod}_{\text{Volitional}}$ | 11. $\text{Mod}_{\text{Obligation}}$ | 12. $\text{Mod}_{\text{Ability/permission}}$ |
| 13. $\text{Asp}_{\text{Habitual}}$ | 14. $\text{Asp}_{\text{Repetitive(I)}}$ | 15. $\text{Asp}_{\text{Frequentative(I)}}$ | 16. $\text{Asp}_{\text{Celerative(I)}}$ |
| 17. T(Anterior) | 18. $\text{Asp}_{\text{Terminative}}$ | 19. $\text{Asp}_{\text{Continuative}}$ | 20. $\text{Asp}_{\text{Perfect(?)}}$ |
| 21. $\text{Asp}_{\text{Retrospective}}$ | 22. $\text{Asp}_{\text{Proximative}}$ | 23. $\text{Asp}_{\text{Durative}}$ | 24. $\text{Asp}_{\text{Generic/progressive}}$ |
| 25. $\text{Asp}_{\text{Prospective}}$ | 26. $\text{Asp}_{\text{SgCompletive(I)}}$ | 27. $\text{Asp}_{\text{PlCompletive}}$ | 28. $\text{Asp}_{\text{Voice}}$ |
| 29. $\text{Asp}_{\text{Celerative(II)}}$ | 30. $\text{Asp}_{\text{SgCompletive(II)}}$ | 31. $\text{Asp}_{\text{Repetitive(II)}}$ | 32. $\text{Asp}_{\text{Frequentative(II)}}$ |

that all sentences contain a structure with all these functional heads. The specifier positions of these heads can be occupied by adverbs or remain empty. Cinque claims that these functional heads and the corresponding structures form part of Universal Grammar, that is, knowledge of these structures is innate (page 107).[9] Laenzlinger (2004) follows Cinque in proposing this sequence of functional heads for German. He also follows Kayne (1994), who assumes that all syntactic structures have the order specifier head complement cross-linguistically, even if the surface order of the constituents seems to contradict this.

The constituent orders that are visible in the end are derived by leftward-movement.[10] Figure 4.20 on the next page shows the analysis of a verb-final clause where the func-

[9] Table 4.1 shows only the functional heads in the clausal domain. Cinque (1994: 96, 99) also accounts for the order of adjectives with a cascade of projections: Quality, Size, Shape, Color, Nationality. These categories and their ordering are also assumed to belong to UG (p. 100).

Cinque (1994: 96) claims that a maximum of seven attributive adjectives are possible and explains this with the fact that there are a limited number of functional projections in the nominal domain. As was shown on page 67, with a fitting context it is possible to use several adjectives of the same kind, which is why some of Cinque's functional projections would have to be subject to iteration.

[10] This also counts for extraposition, that is, the movement of constituents into the postfield in German. Whereas this would normally be analyzed as rightward-movement, Kayne (1994: Chapter 9) analyzes it as movement of everything else to the left. Kayne assumes that (i.b) is derived from (i.a) by moving part of the NP:

(i) a. just walked into the room [$_{\text{NP}}$ someone who we don't know].
b. Someone$_i$ just walked into the room [$_{\text{NP}}$ $\_i$ who we don't know].

(i.a) must have to be some kind of derived intermediate representation, otherwise English would not be SV(O) underlyingly but rather V(O)S. (i.a) is therefore derived from (ii) by fronting the VP *just walked into the room*.

(ii) Someone who we don't know just walked into the room

Such analyses have the downside that they cannot be easily combined with performance models (see Chapter 15).

tional adverbial heads have been omitted.[11] Subjects and objects are generated as argu-

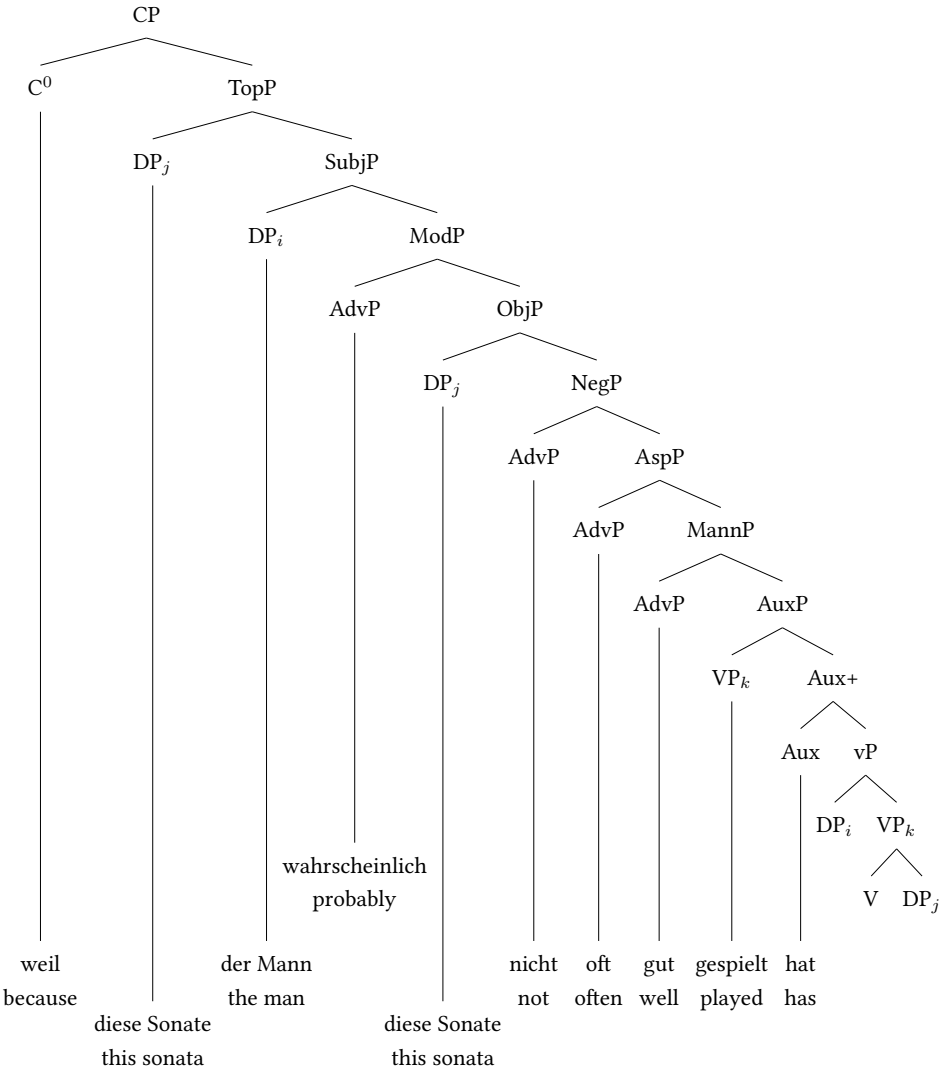



Figure 4.20: Analysis of sentence structure with leftward remnant movement and functional heads following Laenzlinger (2004: 224)

ments inside of vP and VP, respectively. The subject is moved to the specifier of the

[11] These structures do not correspond to $\overline{\text{X}}$ theory as it was presented in Section 2.5. In some cases, heads have been combined with complements to form an XP rather than an X′. For more on $\overline{\text{X}}$ theory in the Minimalist Program, see Section 4.6.3.

subject phrase and the object is moved to the specifier of the object phrase. The verbal projection ($VP_k$) is moved in front of the auxiliary into the specifier position of the phrase containing the auxiliary. The only function of SubjP and ObjP is to provide a landing site for the respective movements. For a sentence in which the object precedes the subject, Laenzlinger assumes that the object moves to the specifier of a topic phrase. Figure 4.20 contains only a ModP and an AspP, although Laenzlinger assumes that all the heads proposed by Cinque are present in the structure of all German clauses. For ditransitive verbs, Laenzlinger assumes multiple object phrases (page 230). A similar analysis with movement of object and subject from verb-initial VPs to Agr positions was suggested by Zwart (1994) for Dutch.

For general criticism of Kayne's model, see Haider (2000). Haider shows that a Kayne-like theory makes incorrect predictions for German (for instance regarding the position of selected adverbials and secondary predicates and regarding verbal complex formation) and therefore fails to live up to its billing as a theory which can explain all languages. Haider (1997a: Section 4) has shown that the assumption of an empty Neg head, as assumed by Pollock (1989), Haegeman (1995) and others, leads to problems. See Bobaljik (1999) for problems with the argumentation for Cinque's cascade of adverb-projections.

Furthermore, it has to be pointed out that SubjP and ObjP, TraP (Transitive Phrase) and IntraP (Intransitive Phrase) (Karimi-Doostan 2005: 1745) and TopP (topic phrase), DistP (quantifier phrase), AspP (aspect phrase) (Kiss 2003: 22; Karimi 2005: 35), PathP and PlaceP (Svenononius 2004: 246) encode information about grammatical function, valence, information structure and semantics in the category symbols.[12] In a sense, this is a misuse of category symbols, but such a misuse of information structural and semantic categories is necessary since syntax, semantics, and information structure are tightly connected and since it is assumed that the semantics interprets the syntax, that is, it is assumed that semantics comes after syntax (see Figure 3.2 and Figure 4.1). By using semantically and pragmatically relevant categories in syntax, there is no longer a clean distinction between the levels of morphology, syntax, semantics and pragmatics: everything has been 'syntactified'. Felix Bildhauer (p. c. 2012) has pointed out to me that approaches which assume a cascade of functional projections where the individual aspects of meaning are represented by nodes are actually very close to phrasal approaches in Construction Grammar (see Adger 2013: 470 also for a similar view). One simply lists configurations and these are assigned a meaning (or features which are interpreted post-syntactically. See Cinque & Rizzi (2010: 62) for the interpretation of TopP, for example).

#### 4.6.1.2 Feature checking in specifier positions

If one takes the theory of feature checking in Specifier-Head relations to its logical conclusion, then one arrives at an analysis such as the one suggested by Radford (1997: 452). Radford assumes that prepositions are embedded in an Agreement Phrase in addition to the structure in (23), which is usually assumed, and that the preposition adjoins to

[12] For further examples and references, see Newmeyer (2004a: 194; 2005: 82). Newmeyer references also works which stipulate a projection for each semantic role. E. g. agent Reciprocal, Benefactive phrase, Instrumental, Causative, Comitative, and Reversive Phrase.

the head of the Agreement Phrase and the argument of the preposition is moved to the specifier position of the Agreement Phrase.

(23) [$_{PP}$ P DP ]

The problem here is that the object now precedes the preposition. In order to rectify this, Radford assumes a functional projection p (read *little p*) with an empty head to which the preposition then adjoins. This analysis is shown in Figure 4.21. This machinery is

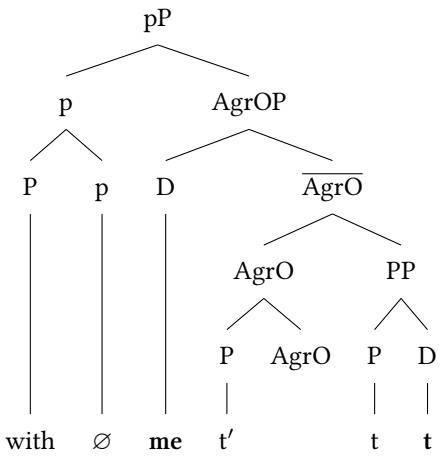



Figure 4.21: PP analysis following Radford with case assignment in specifier position and little p

only necessary in order to retain the assumption that feature checking takes place in specifier-head relations. If one were to allow the preposition to determine the case of its object locally, then all this theoretical apparatus would not be necessary and it would be possible to retain the well-established structure in (23).

Sternefeld (2006: 549–550) is critical of this analysis and compares it to Swiss cheese (being full of holes). The comparison to Swiss cheese is perhaps even too positive since, unlike Swiss cheese, the ratio of substance to holes in the analysis is extreme (2 words vs. 5 empty elements). We have already seen an analysis of noun phrases on page 72, where the structure of an NP, which only consisted of an adjective *klugen* 'clever', contained more empty elements than overt ones. The difference to the PP analysis discussed here is that empty elements are only postulated in positions where overt determiners and nouns actually occur. The little p projection, on the other hand, is motivated entirely theory-internally. There is no theory-external motivation for any of the additional assumptions made for the analysis in Figure 4.21 (see Sternefeld 2006: 549–550).

A variant of this analysis has been proposed by Hornstein, Nunes & Grohmann (2005: 124). The authors do without little p, which makes the structure less complex. They assume the structure in (24), which corresponds to the AgrOP-subtree in Figure 4.21.

(24) $[_{\text{AgrP}}\ \text{DP}_k\ [_{\text{Agr}'}\ \text{P}_i\text{+Agr}\ [_{\text{PP}}\ \text{t}_i\ \text{t}_k\ ]]]$

The authors assume that the movement of the DP to SpecAgrP happens invisibly, that is, covert. This solves Radford's problem and makes the assumption of pP redundant.

The authors motivate this analysis by pointing out agreement phenomena in Hungarian: Hungarian postpositions agree with the preceding noun phrase in person and number. That is, the authors argue that English prepositional and Hungarian postpositional phrases have the same structure derived by movement, albeit the movement is covert in English.

In this way, it is possible to reduce the number and complexity of basic operations and, in this sense, the analysis is minimal. These structures are, however, still incredibly complex. No other kind of theory discussed in this book needs the amount of inflated structure to analyze the combination of a preposition with a noun phrase. The structure in (24) cannot be motivated by reference to data from English and it is therefore impossible to acquire it from the linguistic input. A theory which assumes these kind of structures would have to postulate a Universal Grammar with the information that features can only be checked in (certain) specifier positions (see Chapters 13 and 16 for more on Universal Grammar and language acquisition).

#### 4.6.1.3 Locality of selection and functional projections

Another problem arises from the use of functional heads to encode linear order. In the classic CP/IP-system and all other theories discussed here, a category stands for a class of objects with the same distribution, that is, NP (or DP) stands for pronouns and complex noun phrases. Heads select phrases with a certain category. In the CP/IP-system, I selects a VP and an NP, whereas C selects an IP. In newer analyses, this kind of selectional mechanism does not work as easily. Since movement has taken place in (25b), we are dealing with a TopP or FocP in *das Buch dem Mann zu geben* 'the book the man to give'. Therefore, *um* cannot simply select an non-finite IP, but rather has to disjunctively be able to select a TopP, FocP or IP. It has to be ensured that TopPs and FocPs are marked with regard to the form of the verb contained inside them, since *um* can only be combined with *zu*-infinitives.

(25) a. um dem Mann das Buch zu geben
for the man the book to give
'to give the man the book'

b. um das Buch dem Mann zu geben
for the book the man to give
'to give the book to the man'

The category system, selectional mechanisms and projection of features would therefore have to be made considerably more complicated when compared to a system which simply base generates the orders or a system in which a constituent is moved out of the IP, thereby creating a new IP.

Proposals that follow Cinque (1999) are problematic for similar reasons: Cinque assumes the category AdverbP for the combination of an adverb and a VP. There is an empty functional head, which takes the verbal projection as its complement and the adverb surfaces in the specifier of this projection. In these systems, adverb phrases have to pass on inflectional properties of the verb since verbs with particular inflectional properties (finiteness, infinitives with *zu*, infinitives without *zu*, participles) have to be selected by higher heads (see page 183 and Section 9.1.4). There is of course the alternative to use Agree for this, but then all selection would be nonlocal and after all selection is not agreement. For further, more serious problems with this analysis like modification of adverbs by adverbs in connection with partial fronting and restrictions on non-phrasality of preverbal adverbials in English, see Haider (1997a: Section 5).

A special case of the adverb problem is the negation problem: Ernst (1992) studied the syntax of negation more carefully and pointed out that negation can attach to several different verbal projections (26a,b), to adjectives (26c) and adverbs (26d).

(26) a. Ken could not have heard the news.
b. Ken could have not heard the news.
c. a [not unapproachable] figure
d. [Not always] has she seasoned the meat.

If all of these projections are simply NegPs without any further properties (about verb form, adjective part of speech, adverb part of speech), it would be impossible to account for their different syntactic distributions. Negation is clearly just a special case of the more general problem, since adverbs may attach to adjectives forming adjectival phrases in the traditional sense and not adverb phrases in Chinque's sense. For instance, the adverb *oft* 'often' in (27) modifies *lachender* 'laughing' forming the adjectival phrase *oft lachender*, which behaves like the unmodified adjectival participle *lachender*: it modifies *Mann* 'man' and it precedes it.

(27) a. ein lachender Mann
a laughing man
'a laughing man'
b. ein oft lachender Mann
a often laughing man
'a man that laughs often'

Of course one could imagine solutions to the last three problems that use the Agree relation to enforce selectional constraints nonlocally, but such accounts would violate locality of selection (see Ernst 1992: 110 and the discussion in Section 18.2 of this book) and would be much more complicated than accounts that assume a direct selection of dependents.

Related to the locality issues that were discussed in the previous paragraph is the assumption of special functional projections for the placement of clitics: if one uses SpeakerP so that a clitic for first person singular can be moved to the correct specifier

positions and a HearerP so that the clitic for second person can be moved to the correct position (Poletto 2000: 31), then what one has are special projections which need to encode in addition all features that are relevant for clauses (alternatively one could of course assume nonlocal Agree to be responsible for distributional facts). In addition to these features, the category labels contain information that allows higher heads to select clauses containing clitics. In other approaches and earlier variants of transformational grammar, selection was assumed to be strictly local so that higher heads only have access to those properties of embedded categories that are directly relevant for selection (Abraham 2005: 223; Sag 2007) and not information about whether an argument of a head within the clause is the speaker or the hearer or whether some arguments in the clause are realized as clitics. Locality will be discussed further in Section 18.2.

#### 4.6.1.4 Feature-driven movement

Finally, there is a conceptual problem with feature-driven movement, which has been pointed out by Gisbert Fanselow: Frey (2004a: 27) assumes a KontrP (contrastive phrase) and Frey (2004b) a TopP (topic phrase) (see Rizzi (1997) for TopP and FocP (focus phrase) in Italian and Haftka (1995), Grewendorf (2002: 85, 240); 2009, Abraham (2003: 19), Laenzlinger (2004: 224) and Hinterhölzel (2004: 18) for analyses of German with TopP and/or FocP). Constituents have to move to the specifier of these functional heads depending on their information structural status. Fanselow (2003a) has shown that such movement-based theories for the ordering of elements in the middle field are not compatible with current assumptions of the Minimalist Program. The reason for this is that sometimes movement takes place in order to create space for other elements (altruistic movement). If the information structure of a sentence requires that the closest object to a verb is neither focused nor part of the focus, then the object closest to the verb should not receive the main stress in the clause. This can be achieved by deaccentuation, that is, by moving the accent to another constituent or even, as shown in (28b), by moving the object to a different position from the one in which it receives structural stress.

(28) a. dass die Polizei gestern Linguisten verhaftete
that the police yesterday linguists arrested
'that the police arrested linguists yesterday'

b. dass die Polizei Linguisten gestern verhaftete
that the police linguists yesterday arrested
'that the police arrested linguists yesterday'

In Spanish, partial focus can be achieved not by special intonation, but rather only by altruistic movement in order to move the object out of the focus. See also Bildhauer & Cook (2010: p. 72) for a discussion of 'altruistic' multiple frontings in German.

It is therefore not possible to assume that elements are moved to a particular position in the tree in order to check some feature motivated by information structural properties. Since feature checking is a prerequisite for movement in current minimalist theory, one would have to postulate a special feature, which only has the function of triggering

altruistic movement. Fanselow (2003a: Section 4; 2006: 8) has also shown that the ordering constraints that one assumes for topic, focus and sentence adverbs can be adequately described by a theory which assumes firstly, that arguments are combined (in minimalist terminology: *merged*) with their head one after the other and secondly, that adjuncts can be adjoined to any projection level. The position of sentence adverbs directly before the focused portion of the sentence receives a semantic explanation: since sentence adverbs behave like focus-sensitive operators, they have to directly precede elements that they refer to. It follows from this that elements which do not belong to the focus of an utterance (topics) have to occur in front of the sentence adverb. It is therefore not necessary to assume a special topic position to explain local reorderings in the middle field. This analysis is also pursued in LFG and HPSG. The respective analyses are discussed in more detail in the corresponding chapters.

### 4.6.2 Labeling

In the Minimalist Program, Chomsky tries to keep combinatorial operations and mechanisms as simple as possible. He motivates this with the assumption that the existence of a UG with less language-specific knowledge is more plausible from a evolutionary point of view than a UG which contains a high degree of language-specific knowledge (Chomsky 2008: 135).

For this reason, he removes the projection levels of $\overline{\text{X}}$ theory, traces, indices and "similar descriptive technology" (Chomsky 2008: 138). All that remains is Merge and Move, that is, Internal and External Merge. Internal and External Merge combine two syntactic objects $\alpha$ and $\beta$ into a larger syntactic object which is represented as a set $\{ \alpha, \beta \}$. $\alpha$ and $\beta$ can be either lexical items or internally complex syntactic objects. Internal Merge moves a part of an object to its periphery.[13] The result of internally merging an element is a set $\{ \alpha, \beta \}$ where $\alpha$ was a part of $\beta$. External Merge also produces a set with two elements. However, two independent objects are merged. The objects that are created by Merge have a certain category (a set of features). For instance, if one combines the elements $\alpha$ and $\beta$, one gets $\{ \text{l}, \{ \alpha, \beta \} \}$, where l is the category of the resulting object. This category is also called a *label*. Since it is assumed that all constituents are headed, the category that is assigned to $\{ \alpha, \beta \}$ has to be either the category of $\alpha$ or the category of $\beta$. Chomsky (2008: 145) discusses the following two rules for the determination of the label of a set.

(29) a. In $\{ \text{H}, \alpha \}$, H an LI, H is the label.
b. If $\alpha$ is internally merged to $\beta$, forming $\{ \alpha, \beta \}$ then the label of $\beta$ is the label of $\{ \alpha, \beta \}$.

As Chomsky notes, these rules are not unproblematic since the label is not uniquely determined in all cases. An example is the combination of two lexical elements. If both

[13] To be more specific, part of a syntactic object is copied and the copy is placed at the edge of the entire object. The original of this copy is no longer relevant for pronunciation (*Copy Theory of Movement*).

H and $\alpha$ in (29a) are lexical items (LI), then both H and $\alpha$ can be the label of the resulting structure. Chomsky notices that this could result in deviant structures, but claims that this concern is unproblematic and ignores it. Chomsky offered a treatment of the combination of two lexical items in 2013. The solution to the problem is to assume that all combinations of lexical elements consist of a functional element and a root (Marantz 1997; Borer 2005). Roots are not considered as labels per definition[14] and hence the category of the functional element determines the category of the combination (Chomsky 2013: 47). Such an analysis can only be rejected: the goal of the Minimalist Program is to simplify the theoretical proposals to such an extent that the models of language acquisition and language evolution become plausible, but in order to simplify basic concepts it is stipulated that a noun cannot simply be a noun but needs a functional element to tell the noun what category it has. Given that the whole point of Chomsky's Bare Phrase Structure (Chomsky 1995a) was the elimination of the unary branching structures in $\overline{\text{X}}$ theory, it is unclear why they are reintroduced now through the backdoor, only more complex with an additional empty element.[15] Theories like Categorial Grammar and HPSG can combine lexical items directly without assuming any auxiliary projections or empty elements. See also Rauh (2013) for a comparison of the treatment of syntactic categories in earlier versions of Transformational Grammar, HPSG, Construction Grammar, Role and Reference Grammar and root-based Neo-Constructivist proposals like the one assumed by Chomsky (2013). Rauh concludes that the direct connection of syntactic and semantic information is needed and that the Neo-Constructivism of Marantz and Borer has to be rejected. For further criticism of Neo-Constructivist approaches see Wechsler (2008a) and Müller & Wechsler (2014a: Sections 6.1 and 7).

The combination of a pronoun with a verbal projection poses a problem that is related to what has been said above. In the analysis of *He left*, the pronoun *he* is a lexical element and hence would be responsible for the label of *He left*, since *left* is an internally complex verbal projection in Minimalist theories. The result would be a nominal label rather than a verbal one. To circumvent this problem, Chomsky (2013: 46) assumes that *he* has a complex internal structure: 'perhaps D-pro', that is, *he* is (perhaps) composed out of an invisible determiner and a pronoun.

The case in which two non-LIs are externally merged (for instance a nominal and a verbal phrase) is not discussed in Chomsky (2008). Chomsky (2013: 43–44) suggests that a phrase XP is irrelevant for the labeling of { XP, YP } if XP is moved (or rather copied in the Copy Theory of Movement) in a further step. Chomsky assumes that one of two phrases in an { XP, YP } combination has to move, since otherwise labeling would be

[14] Another category that is excluded as label per definition is *Conj*, which stands for conjunction (Chomsky 2013: 45–46). This is a stipulation that is needed to get coordination to work. See below.

[15] The old $\overline{\text{X}}$ rule in (i.a) corresponds to the binary combination in (i.b).

(i) a. $\text{N}' \rightarrow \text{N}$
b. $\text{N} \rightarrow$ N-func root

In (i.a) a lexical noun is projected to an $\text{N}'$ and in (i.b) a root is combined with a functional nominal head into a nominal category.

impossible (p. 12).[16] The following coordination example will illustrate this: Chomsky assumes that the expression *Z and W* is analyzed as follows: first, Z and W are merged. This expression is combined with Conj (30a) and in the next step Z is raised (30b).

(30) a. [$_\alpha$ Conj [$_\beta$ Z W]]
b. [$_\gamma$ Z [$_\alpha$ Conj [$_\beta$ Z W]]

Since Z in $\beta$ is only a copy, it does not count for labeling and $\beta$ can get the label of W. It is stipulated for the combination of Z and $\alpha$ that Conj cannot be the label and hence the label of the complete structure is Z.[17]

A special case that is discussed by Chomsky is the Internal Merge of an LI $\alpha$ with a non LI $\beta$. According to rule (29a) the label would be $\alpha$. According to (29b), the label would be $\beta$ (see also Donati (2006)). Chomsky discusses the combination of the pronoun *what* with *you wrote* as an example.

(31) what [ C [you wrote *t*]]

If the label is determined according to (29b), one then has a syntactic object that would be called a CP in the GB framework; since this CP is, moreover, interrogative, it can function as the complement of *wonder* as in (32a). If the label is determined according to (29a), one gets an object that can function as the accusative object of *read* in (32b), that is, something that corresponds to a DP in GB terminology.

(32) a. I wonder what you wrote.
b. I read what you wrote.

*what you wrote* in (32b) is a so-called free relative clause.

Chomsky's approach to free relative clauses is interesting but is unable to describe the phenomenon in full breadth. The problem is that the phrase that contains the relative pronoun may be complex (contrary to Donati's claims, see also Citko (2008: 930–932)).[18]

[16] His explanation is contradictory: on p. 11 Chomsky assumes that a label of a combination of two entities with the same category is this category. But in his treatment of coordination, he assumes that one of the conjuncts has to be raised, since otherwise the complete structure could not be labeled.

[17] As Bob Borsley (p.c. 2013) pointed out to me, this makes wrong predictions for coordinations of two singular noun phrases with *and*, since the result of the coordination is a plural NP and not a singular one like the first conjunct. Theories like HPSG can capture this by grouping features in bundles that can be shared in coordinated structures (syntactic features and nonlocal features, see Pollard & Sag (1994: 202)).

Furthermore the whole account cannot explain why (i.b) is ruled out.

(i) a. both Kim and Lee
b. * both Kim or Lee

The information about the conjunction has to be part of the representation for *or Lee* in order to be able to contrast it with *and Lee*.

A further problem is that the label of $\alpha$ should be the label of W since Conj does not count for label determination. This would lead to a situation in which we have to choose between Z and W to determine the label of $\gamma$. Following Chomsky's logic, either Z or W would have to move on to make it possible to label $\gamma$. Chomsky mentions this problem in footnote 40, but does not provide a solution.

[18] Chomsky (2013: 47) admits that there are many open questions as far as the labeling in free relative clauses is concerned and hence admits that there remain many open questions with labeling as such.

(33) provides an English example from Bresnan & Grimshaw (1978: 333). German examples from Bausewein (1990: 155) and Müller (1999b: 78) are given in (34).

(33) I'll read [whichever book] you give me.

(34) a. Ihr könnt beginnen, [mit *wem*] ihr wollt.[19]
you can start with whom you want
'You can start with whoever you like.'

b. [*Wessen* Birne] noch halbwegs in der Fassung steckt, pflegt solcherlei
whose bulb/head yet halfway in the socket is uses such
Erloschene zu meiden;[20]
extinct to avoid
'Those who still have their wits half way about them tend to avoid such vacant characters;'

c. [*Wessen* Schuhe] „danach" besprenkelt sind, hat keinen Baum gefunden und
whose shoes after.that speckled are has no tree found and
war nicht zu einem Bogen in der Lage.[21]
was not to a bow in the position
'Those whose shoes are spattered afterwards couldn't find a tree and were incapable of peeing in an arc.'

Since *wessen Schuhe* 'whose shoes' is not a lexical item, rule (29b) has to be applied, provided no additional rules are assumed to deal with such cases. This means that the whole free relative clause *wessen Schuhe danach besprenkelt sind* is labeled as CP. For the free relatives in (33) and (34) the labeling as a CP is an unwanted result, since they function as subjects or objects of the matrix predicates and hence should be labelled DP. However, since *wessen Schuhe* is a complex phrase and not a lexical item, (29a) does not apply and hence there is no analysis of the free relative clause as a DP. Therefore, it seems one must return to something like the GB analysis proposed by Groos & van Riemsdijk (1981), at least for the German examples. Gross and van Riemsdijk assume that free relatives consist of an empty noun that is modified by the relative clause like a normal noun. In such an approach, the complexity of the relative phrase is irrelevant. It is only the empty head that is relevant for labeling the whole phrase.[22] However, once empty heads are countenanced in the analysis, the application of (29a) to (31) is

[19] Bausewein (1990: 155).
[20] Thomas Gsella, taz, 12.02.1997, p. 20.
[21] taz, taz mag, 08./09.08.1998, p. XII.
[22] Assuming an empty head is problematic since it may be used as an argument only in those cases in which it is modified by an adjunct, namely the relative clause (Müller 1999b: 97). See also Ott (2011: 187) for a later rediscovery of this problem. It can be solved in HPSG by assuming a unary projection that projects the appropriate category from a relative clause. I also use the unary projection to analyze so-called *nonmatching* free relative clauses (Müller 1999b). In constructions with nonmatching free relative clauses, the relative clause fills an argument slot that does not correspond to the properties of the relative phrase (Bausewein 1990). Bausewein discusses the following example, in which the relative phrase is a PP but the free relative fills the accusative slot of *kocht* 'cooks'.

undesirable since the application would result in two analyses for (32b): one with the empty nominal head and one in which (31) is labeled as NP directly. One might argue that in the case of several possible derivations, the most economical one wins, but the assumption of transderivational constraints leads to undesired consequences (Pullum 2013: Section 5).

Chomsky (2013) abandons the labeling condition in (29b) and replaces it with general labeling rules that hold for both internal and external Merge of two phrases. He distinguishes two cases. In the first case, labeling becomes possible since one of the two phrases of the set { XP, YP } is moved away. This case was already discussed above. Chomsky writes about the other case: *X and Y are identical in a relevant respect, providing the same label, which can be taken as the label of the SO* (p. 11). He sketches an analysis of interrogative clauses on p. 13 in which the interrogative phrase has a Q feature and the remaining sentence from which the Q phrase was extracted has a Q feature as well. Since the two constituents share this property, the label of the complete clause will be Q. This kind of labeling will "perhaps" also be used for labeling normal sentences consisting of a subject and a verb phrase agreeing in person and number. These features would be responsible for the label of the sentence. The exact details are not worked out, but almost certainly will be more complex than (29b).

A property that is inherent in both Chomsky (2005) and Chomsky (2013) is that the label is exclusively determined from one of the merged objects. As Bob Borsley pointed out to me, this is problematic for interrogative/relative phrases like (35).

(35) with whom

The phrase in (35) is both a prepositional phrase (because the first word is a preposition) and an interrogative/relative phrase (because the second word is an interroga-

(i) Sie kocht, worauf sie Appetit hat.
she cooks where.on she appetite has
'She cooks what she feels like eating.'

See Müller (1999b: 60–62) for corpus examples.

Minimalist theories do not employ unary projections. Ott (2011) develops an analysis in which the category of the relative phrase is projected, but he does not have a solution for nonmatching free relative clauses (p. 187). The same is true for Citko's analysis, in which an internally merged XP can provide the label.

Many other proposals for labeling or, rather, non-labeling exist. For instance, some Minimalists want to eliminate labeling altogether and argue for a label-free syntax. As was pointed out by Osborne, Putnam & Groß (2011), such analyses bring Minimalism closer to Dependency Grammar. It is unclear how any of these models could deal with non-matching free relative clauses. Groß & Osborne (2009: Section 5.3.3) provide an analysis of free relatives in their version of Dependency Grammar, but deny the existence of nonmatching ones (p. 78). They suggest an analysis in which the relative phrase is the root/label of the free relative clause and hence they have the same problem as Minimalist proposals have with non-matching free relative clauses. As Groß & Osborne (2009: 73) and Osborne et al. (2011: 327) state: empty heads are usually not assumed in (their version of) Dependency Grammar. Neither are unary branching projections. This seems to make it impossible to state that free relative clauses with a relative phrase YP can function as XP, provided XP is a category that is higher in the obliqueness hierarchy of Keenan & Comrie (1977), a generalization that was discovered by Bausewein (1990) (see also Müller 1999b: 60–62 and Vogel 2001: 4). In order to be able to express the relevant facts, an element or a label has to exist that is different from the label of *worauf* in (i).

tive/relative word). So, what is needed for the correct labeling of PPs like the one in (35) is a well-defined way of percolating different properties from daughters to the mother node.[23]

Summarizing, one can say that labeling, which was introduced to simplify the theory and reduce the amount of language specific innate knowledge that has to be assumed, can only be made to function with a considerable amount of stipulations. For instance, the combination of lexical elements requires the assumption of empty functional heads, whose only purpose is determining the syntactic category of a certain lexical element. If this corresponded to linguistic reality, knowledge about labeling, the respective functional categories, and information about those categories that have to be ignored for the labeling would have to be part of innate language specific knowledge and nothing would be gained. One would be left with bizarre analyses with an enormous degree of complexity without having made progress in the Minimalist direction. Furthermore, there are empirical problems and a large number of unsolved cases.

The conclusion is that the label of a binary combination should not be determined in the ways suggested by Chomsky (2008, 2013). An alternative option for computing the label is to use the functor of a functor argument structure as the label (Berwick & Epstein 1995: 145). This is the approach taken by Categorial Grammar (Ajdukiewicz 1935; Steedman 2000) and in Stabler's Minimalist Grammars (2011).[24] Stabler's formalization of Merge will be discussed in Section 4.6.4.

### 4.6.3 Specifiers, complements, and the remains of $\overline{\text{X}}$ theory

Chomsky (2008: 146) assumes that every head has exactly one complement but an arbitrary number of specifiers. In standard $\overline{\text{X}}$ theory, the restriction that there can be at most one complement followed from the general $\overline{\text{X}}$ schema and the assumption that structures are at most binary branching: in standard $\overline{\text{X}}$ theory a lexical head was combined with

[23] HPSG solves this problem by distinguishing head features including part of speech information and nonlocal features containing information about extraction and interrogative/relative elements. Head features are projected from the head, the nonlocal features of a mother node are the union of the nonlocal features of the daughters minus those that are bound off by certain heads or in certain configurations.

Citko (2008: 926) suggests an analysis in which both daughters can contribute to the mother node. The result is a complex label like { P, { D, N } }. This is a highly complex data structure and Citko does not provide any information on how the relevant information that it contains is accessed. Is an object with the label { P, { D, N } } a P, a D or an N? One could say that P has priority since it is in the least embedded set, but D and N are in one set. What about conflicting features? How does a preposition that selects for a DP decide whether { D, N } is a D or an N? In any case it is clear that a formalization will involve recursive relations that dig out elements of subsets in order to access their features. This adds to the overall complexity of the proposal and is clearly dispreferred over the HPSG solution, which uses one part of speech value per linguistic object.

[24] For the Categorial Grammar approach to work, it is necessary to assign the category x/x to an adjunct, where x stands for the category of the head to which the adjunct attaches. For instance, an adjective combines with a nominal object to form a nominal object. Therefore its category is n/n rather than adj.

Similarly, Stabler's approach does not extend to adjuncts unless he is willing to assign the category noun to attributive adjectives. One way out of this problem is to assume a special combination operation for adjuncts and their heads (see Frey & Gärtner 2002: Section 3.2). Such a combination operation is equivalent to the Head-Adjunct Schema of HPSG.

all its complements to form an X′. If there are at most two daughters in a phrase, it follows that there can be only one complement (Sentences with ditransitive verbs have been analyzed with an empty head licensing an additional argument; see Larson (1988) for the suggestion of an empty verbal head and Müller & Wechsler (2014a: Sections 6.1 and 7) for a critical assessment of approaches involving little *v*). In standard $\overline{\text{X}}$ theory there was just one specifier. This restriction has now been abandoned. Chomsky writes that the distinction between specifier and complement can now be derived from the order in which elements are merged with their head: elements that are *first-merged* are complements and all others – those which are *later-merged* – are specifiers.

Such an approach is problematic for sentences with mono-valent verbs: according to Chomsky's proposal, subjects of mono-valent verbs would not be specifiers but complements.[25] This problem will be discussed in more detail in Section 4.6.4.

Apart from this, theories assuming that syntactic objects merged with word groups are specifiers do not allow for analyses in which two lexical verbs are directly coordinated as in (36):[26]

(36) He [knows and loves] this record.

For example, in an analysis suggested by Steedman (1991: 264), *and* (being the head) is first merged with *loves* and then the result is merged with *knows*. The result of this combination is a complex object that has the same syntactic properties as the combined parts: the result is a complex verb that needs a subject and an object. After the combination of the conjunction with the two verbs, the result has to be combined with *this record* and *he*. *this record* behaves in all relevant respects like a complement. Following Chomsky's definition, however, it should be a specifier, since it is combined with

[25] Pauline Jacobson (p.c. 2013) pointed out that the problem with intransitive verbs could be solved by assuming that the last-merged element is the specifier and all non-last-merged elements are complements. This would solve the problems with intransitive verbs and with the coordination of verbs in (36) but it would not solve the problem of coordination in head-final languages as in (39). Furthermore, current Minimalist approaches make use of multiple specifiers and this would be incompatible with the Jacobsonian proposal unless one would be willing to state more complicated restrictions on the status of non-first-merged elements.

[26] Chomsky (2013: 46) suggests the coordination analysis in (30): according to this analysis, the verbs would be merged directly and one of the verbs would be moved around the conjunction in a later step of the derivation. As was mentioned in the previous section, such analyses do not contribute to the goal of making minimal assumptions about innate language specific knowledge since it is absolutely unclear how such an analysis of coordination would be acquired by language learners. Hence, I will not consider this coordination analysis here.

Another innovation of Chomsky's 2013 paper is that he eliminates the concept of specifier. He writes in footnote 27 on page 43: *There is a large and instructive literature on problems with Specifiers, but if the reasoning here is correct, they do not exist and the problems are unformulable.* This is correct, but this also means that everything that was explained with reference to the notion of specifier in the Minimalist framework until now does not have an explanation any longer. If one follows Chomsky's suggestion, a large part of the linguistic research of the past years becomes worthless and has to be redone.

Chomsky did not commit himself to a particular view on linearization in his earlier work, but somehow one has to ensure that the entities that were called specifier are realized in a position in which constituents are realized that used to be called specifier. This means that the following remarks will be relevant even under current Chomskyan assumptions.

the third application of Merge. The consequences are unclear. Chomsky assumes that Merge does not specify constituent order. According to him the linearization happens at the level of Phonological Form (PF). The restrictions that hold there are not described in his recent papers. However, if the categorization as complement or specifier plays a role for linearization as in Kayne's work (2011: 2, 12) and in Stabler's proposal (see Section 4.6.4), *this record* would have to be serialized before *knows and loves*, contrary to the facts. This means that a Categorial Grammar-like analysis of coordination is not viable and the only remaining option would seem to assume that *knows* is combined with an object and then two VPs are coordinated. Kayne (1994: 61, 67) follows Wexler & Culicover (1980: 303) in suggesting such an analysis and assumes that the object in the first VP is deleted. However, Borsley (2005: 471) shows that such an analysis makes wrong predictions, since (37a) would be derived from (37b) although these sentences differ in meaning.[27]

(37) a. Hobbs whistled and hummed the same tune.
b. Hobbs whistled the same tune and hummed the same tune.

Since semantic interpretation cannot see processes such as deletion that happen at the level of Phonological Form (Chomsky 1995b: Chapter 3), the differences in meaning cannot be explained by an analysis that deletes material.

In a further variant of the VP coordination analysis, there is a trace that is related to *this record*. This would be a *Right-Node-Raising* analysis. Borsley (2005) has shown that such analyses are problematic. Among the problematic examples that he discusses is the following pair (see also Bresnan 1974: 615).

(38) a. He tried to persuade and convince him.
b. * He tried to persuade, but couldn't convince, him.

The second example is ungrammatical if *him* is not stressed. In contrast, (38a) is well-formed even with unstressed *him*. So, if (38a) were an instance of Right-Node-Raising, the contrast would be unexpected. Borsley therefore excludes a Right-Node-Raising analysis.

The third possibility to analyze sentences like (36) assumes discontinuous constituents and uses material twice: the two VPs *knows this record* and *loves this record* are coordinated with the first VP being discontinuous. (See Crysmann (2001) and Beavers & Sag (2004) for such proposals in the framework of HPSG.) However, discontinuous constituents are not usually assumed in the Minimalist framework (see for instance Kayne (1994: 67)). Furthermore, Abeillé (2006) showed that there is evidence for structures in which lexical elements are coordinated directly. This means that one needs analyses like the CG analysis discussed above, which would result in the problems with the specifier/complement status just discussed.

[27] See also Bartsch & Vennemann (1972: 102), Jackendoff (1977: 192–193), Dowty (1979: 143), den Besten (1983: 104–105), Klein (1985: 8–9) and Eisenberg (1994b) for similar observations and criticism of similar proposals in earlier versions of Transformational Grammar.

Furthermore, Abeillé has pointed out that NP coordinations in head-final languages like Korean and Japanese present difficulties for Merge-based analyses. (39) shows a Japanese example.

(39) Robin-to Kim
Robin-and Kim
'Kim and Robin'

In the first step *Robin* is merged with *to*. In a second step *Kim* is merged. Since *Kim* is a specifier, one would expect that *Kim* is serialized before the head as it is the case for other specifiers in head-final languages.

Chomsky tries to get rid of the unary branching structures of standard $\overline{\text{X}}$ theory, which were needed to project lexical items like pronouns and determiners into full phrases, referring to work by Muysken (1982b). Muysken used the binary features MIN and MAX to classify syntactic objects as minimal (words or word-like complex objects) or maximal (syntactic objects that stand for complete phrases). Such a feature system can be used to describe pronouns and determiners as [+MIN, +MAX]. Verbs like *give*, however, are classified as [+MIN, −MAX]. They have to project in order to reach the [+MAX]-level. If specifiers and complements are required to be [+MAX], then determiners and pronouns fulfill this requirement without having to project from $\text{X}^0$ via X′ to the XP-level.

In Chomsky's system, the MIN/MAX distinction is captured with respect to the completeness of heads (complete = phrase) and to the property of being a lexical item. However, there is a small but important difference between Muysken's and Chomsky's proposal: the predictions with regard to the coordination data that was discussed above. Within the category system of $\overline{\text{X}}$ theory, it is possible to combine two $\text{X}^0$s to get a new, complex $\text{X}^0$. This new object has basically the same syntactic properties that simple $\text{X}^0$s have (see Jackendoff 1977: 51 and Gazdar, Klein, Pullum & Sag 1985). In Muysken's system, the coordination rule (or the lexical item for the conjunction) can be formulated such that the coordination of two +MIN items is a +MIN item. In Chomsky's system an analogous rule cannot be defined, since the coordination of two lexical items is not a lexical item any longer.

Like Chomsky in his recent Minimalist work, Categorial Grammar (Ajdukiewicz 1935) and HPSG (Pollard and Sag: 1987; 1994: 39–40) do not (strictly) adhere to $\overline{\text{X}}$ theory. Both theories assign the symbol NP to pronouns (for CG see Steedman & Baldridge (2006: p. 615), see Steedman (2000: Section 4.4) for the incorporation of lexical type raising in order to accommodate quantification). The phrase *likes Mary* and the word *sleeps* have the same category in Categorial Grammar (s\np). In both theories it is not necessary to project a noun like *tree* from $\text{N}^0$ to $\overline{\text{N}}$ in order to be able to combine it with a determiner or an adjunct. Determiners and mono-valent verbs in controlled infinitives are not projected from an $\text{X}^0$ level to the XP level in many HPSG analyses, since the valence properties of the respective linguistic objects (an empty SUBCAT or COMPS list) are sufficient to determine their combinatoric potential and hence their distribution (Müller 1996d; Müller 1999a). If the property of being minimal is needed for the description of a phenomenon, the binary feature LEX is used in HPSG (Pollard and Sag: 1987: 172; 1994:

22). However, this feature is not needed for the distinction between specifiers and complements. This distinction is governed by principles that map elements of an argument structure list (ARG-ST) onto valence lists that are the value of the SPECIFIER and the COMPLEMENTS feature (abbreviated as SPR and COMPS respectively).[28] Roughly speaking, the specifier in a verbal projection is the least oblique argument of the verb for configurational languages like English. Since the argument structure list is ordered according to the obliqueness hierarchy of Keenan & Comrie (1977), the first element of this list is the least oblique argument of a verb and this argument is mapped to the SPR list. The element in the SPR list is realized to the left of the verb in SVO languages like English. The elements in the COMPS list are realized to the right of their head. Approaches like the one by Ginzburg & Sag (2000: 34, 364) that assume that head-complement phrases combine a word with its arguments have the same problem with coordinations like (36) since the head of the VP is not a word.[29] However, this restriction for the head can be replaced by one that refers to the LEX feature rather than to the property of being a word or lexical item.

Pollard and Sag as well as Sag and Ginzburg assume flat structures for English. Since one of the daughters is marked as lexical, it follows that the rule does not combine a head with a subset of its complements and then apply a second time to combine the result with further complements. Therefore, a structure like (40a) is excluded, since *gave John* is not a word and hence cannot be used as the head daughter in the rule.

(40) a. [[gave John] a book]
b. [gave John a book]

Instead of (40a), only analyses like (40b) are admitted; that is, the head is combined with all its arguments all in one go. The alternative is to assume binary branching structures (Müller 2015c; Müller & Ørsnes 2015: Section 1.2.2). In such an approach, the head complement schema does not restrict the word/phrase status of the head daughter. The binary branching structures in HPSG correspond to External Merge in the MP.

In the previous two sections, certain shortcomings of Chomsky's labeling definition and problems with the coordination of lexical items were discussed. In the following section, I discuss Stabler's definition of Merge in Minimalist Grammar, which is explicit about labeling and in one version does not have the problems discussed above. I will show that his formalization corresponds rather directly to HPSG representations.

[28] Some authors assume a three-way distinction between subjects, specifiers, and complements.

[29] As mentioned above, a multidomination approach with discontinuous constituents is a possible solution for the analysis of (36) (see Crysmann 2001 and Beavers & Sag 2004). However, the coordination of lexical items has to be possible in principle as Abeillé (2006) has argued. Note also that the HPSG approach to coordination cannot be taken over to the MP. The reason is that the HPSG proposals involve special grammar rules for coordination and MP comes with the claim that there is only Merge. Hence the additional introduction of combinatorial rules is not an option within the MP.

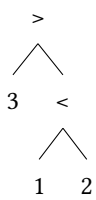
### 4.6.4 Minimalism, Categorial Grammar, and HPSG

In this section, I will relate Minimalism, Categorial Grammar and HPSG to one another. Readers who are not yet familiar with Categorial Grammar and HPSG should skim this section or consult the Chapters 6, 8 and 9 and return here afterwards.

In Section 4.6.2, it was shown that Chomsky's papers leave many crucial details about labeling unspecified. Stabler's work is relatively close to recent Minimalist approaches, but is worked out much more precisely (see also Stabler (2010: 397, 399, 400) on formalization of post GB approaches). Stabler (2001) shows how Kayne's theory of remnant movement can be formalized and implemented. Stabler refers to his particular way of formalizing Minimalist theories as *Minimalist Grammars* (MG). There are a number of interesting results with regard to the weak capacity of Minimalist Grammars and variants thereof (Michaelis 2001). It has been shown, for instance, that the number of possible languages one could create with MGs includes the set of those which can be created by Tree Adjoining Grammars (see Chapter 12). This means that it is possible to assign a greater number of word strings to structures with MGs, however, the structures derived by MGs are not necessarily always the same as the structures created by TAGs. For more on the generative capacity of grammars, see Chapter 17.

Although Stabler's work can be regarded as a formalization of Chomsky's Minimalist ideas, Stabler's approach differs from Chomsky's in certain matters of detail. Stabler assumes that the results of the two Merge operations are not sets but pairs. The head in a pair is marked by a pointer ('<' or '>'). Bracketed expressions like { $\alpha$, { $\alpha$, $\beta$ } } (discussed in Section 4.6.2) are replaced by trees like the one in (41).

(41)




3 1 2

1 is the head in (41), 2 is the complement and 3 the specifier. The pointer points to the part of the structure that contains the head. The daughters in a tree are ordered, that is, 3 is serialized before 1 and 1 before 2.

Stabler (2010: 402) defines External Merge as follows:

(42) em($t_1$[=f], $t_2$[f]) =

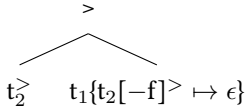

$t_1$ $t_2$ if $t_1$ has exactly 1 node


$t_2$ $t_1$ otherwise

=f is a selection feature and f the corresponding category. When $t_1$[=f] and $t_2$[f] are combined, the result is a tree in which the selection feature of $t_1$ and the respective category feature of $t_2$ are deleted. The upper tree in (42) represents the combination of a (lexical) head with its complement. $t_1$ is positioned before $t_2$. The condition that $t_1$ has to have exactly one node corresponds to Chomsky's assumption that the first Merge is a Merge with a complement and that all further applications of Merge are Merges with specifiers (Chomsky 2008: 146).

Stabler defines Internal Merge as follows:[30]

(43) $\mathrm{im}(t_1[+f]) =$




$t_1$ is a tree with a subtree $t_2$ which has the feature f with the value '−'. This subtree is deleted ($t_2[-f]^{>} \mapsto \epsilon$) and a copy of the deleted subtree without the −f feature ($t_2^{>}$) is positioned in specifier position. The element in specifier position has to be a maximal projection. This requirement is visualized by the raised '>'.

Stabler provides an example derivation for the sentence in (44).

(44) who Marie praises

*praises* is a two-place verb with two =D features. This encodes the selection of two determiner phrases. *who* and *Marie* are two Ds and they fill the object and subject position of the verb. The resulting verbal projection *Marie praises who* is embedded under an empty complementizer which is specified as +WH and hence provides the position for the movement of *who*, which is placed in the specifier position of CP by the application of Internal Merge. The −WH feature of *who* is deleted and the result of the application of Internal Merge is *who Marie praises*.

This analysis has a problem that was pointed out by Stabler himself in unpublished work cited by Veenstra (1998: 124): it makes incorrect predictions in the case of monovalent verbs. If a verb is combined with an NP, the definition of External Merge in (42) treats this NP as a complement[31] and serializes it to the right of the head. Instead of analyses of sentences like (45a) one gets analyses of strings like (45b).

(45) a. Max sleeps.
b. * Sleeps Max.

To solve this problem, Stabler assumes that mono-valent verbs are combined with a nonovert object (see Veenstra (1998: 61, 124) who, quoting Stabler's unpublished work, also adopts this solution). With such an empty object, the resulting structure contains the empty object as a complement. The empty object is serialized to the right of the verb and *Max* is the specifier and hence serialized to the left of the verb as in (46)).

[30] In addition to what is shown in (43), Stabler's definition contains a variant of the *Shortest Move Constraint* (SMC), which is irrelevant for the discussion at hand and hence will be omitted.

[31] Compare also Chomsky's definition of specifier and complement in Section 4.6.3.

(46) Max sleeps _.

Of course, any analysis of this kind is both stipulative and entirely ad hoc, being motivated only by the wish to have uniform structures. Moreover, it exemplifies precisely one of the methodological deficiencies of Transformational Generative Grammar discussed at length by Culicover & Jackendoff (2005: Section 2.1.2): the excessive appeal to uniformity.

An alternative is to assume an empty verbal head that takes *sleeps* as complement and *Max* as subject. Such an analysis is often assumed for ditransitive verbs in Minimalist theories which assume Larsonian verb shells (Larson 1988). Larsonian analyses usually assume that there is an empty verbal head that is called little *v* and that contributes a causative meaning. As was discussed in Section 4.1.4, Adger (2003) adopts a little *v*-based analysis for intransitive verbs. Omitting the TP projection, his analysis is provided in Figure 4.22. Adger argues that the analysis of sentences with unergative verbs involves

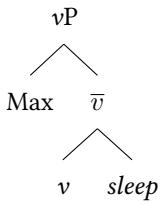



Figure 4.22: Little *v*-based analysis of *Max sleeps*

a little *v* that selects an agent, while the analysis of unaccusative verbs involves a little *v* that does not select an N head. For unaccusatives he assumes that the verb selects a theme. He states that little *v* does not necessarily have a causative meaning but introduces the agent. But note that in the example at hand the subject of *sleep* is neither causing an event nor is it necessarily deliberately doing something. So it is rather an undergoer than an agent. This means that the assumption of the empty *v* head is made for purely theory-internal reasons without any semantic motivation in the case of intransitives. If the causative contribution of little *v* in ditransitive constructions is assumed, this would mean that one needs two little *v*s, one with and one without a causative meaning. In addition to the lack of theory-external motivation for little *v*, there are also empirical problems for such analyses (for instance with coordination data). The reader is referred to Müller & Wechsler (2014a: Sections 6.1 and 7) for further details.

Apart from the two operations that were defined in (42) and (43), there are no other operations in MG.[32] Apart from the problems with mono-valent verbs, this results in the problem that was discussed in Section 4.6.3: there is no analysis with a direct combination of verbs for (36) – repeated here as (47).

(47) He [knows and loves] this record.

[32] For extensions see Frey & Gärtner (2002: Section 3.2).

The reason is that the combination of *knows*, *and* and *loves* consists of three nodes and the Merge of *knows and loves* with *this record* would make *this record* the specifier of the structure. Therefore *this record* would be serialized before *knows and loves*, contrary to the facts. Since the set of languages that can be generated with MGs contains the languages that can be generated with certain TAGs and with Combinatorial Categorial Grammar (Michaelis 2001), the existence of a Categorial Grammar analysis implies that the coordination examples can be derived in MGs somehow. But for linguists, the fact that it is possible to generate a certain string at all (the weak capacity of a grammar) is of less significance. It is the actual structures that are licensed by the grammar that are important (the strong capacity).

#### 4.6.4.1 Directional Minimalist Grammars and Categorial Grammar

Apart from reintroducing $X^0$ categories, the coordination problem can be solved by changing the definition of Merge in a way that allows heads to specify the direction of combination with their arguments: Stabler (2011: p. 635) suggests marking the position of an argument relative to its head together with the selection feature and gives the following redefinition of External Merge.

(48) $em(t_1[\alpha], t_2[x]) = \begin{cases} \begin{array}{c} < \\ t_1 \quad t_2 \end{array} & \text{if } \alpha \text{ is =x} \\ \begin{array}{c} > \\ t_2 \quad t_1 \end{array} & \text{if } \alpha \text{ is x=} \end{cases}$

The position of the equal sign specifies on which side of the head an argument has to be realized. This corresponds to forward and backward Application in Categorial Grammar (see Section 8.1.1). Stabler calls this form of grammar Directional MG (DMG). This variant of MG avoids the problem with mono-valent verbs and the coordination data is unproblematic as well if one assumes that the conjunction is a head with a variable category that selects for elements of the same category to the left and to the right of itself. *know* and *love* would both select an object to the right and a subject to the left and this requirement would be transferred to *knows and loves*.[33] See Steedman (1991: 264) for the details of the CG analysis and Bouma & van Noord (1998: 52) for an earlier HPSG proposal involving directionality features along the lines suggested by Stabler for his DMGs.

[33] Note however, that this transfer makes it necessary to select complex categories, a fact that I overlooked in Müller (2013c). The selection of simplex features vs. complex categories will be discussed in Section 4.6.5.

#### 4.6.4.2 Minimalist Grammars and Head-Driven Phrase Structure Grammar

The notation for marking the head of a structure with '>' and '<' corresponds directly to the HPSG representation of heads. Since HPSG is a sign-based theory, information about all relevant linguistic levels is represented in descriptions (phonology, morphology, syntax, semantics, information structure). (49) gives an example: the lexical item for the word *grammar*.

(49)
$$\begin{bmatrix} \textsc{phon} & \langle\ \textit{'gramər}\ \rangle \\ \textsc{synsem|loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head} & \textit{noun} \\ \textsc{spr} & \langle\ \text{DET}\ \rangle \\ \textit{cat} & \end{bmatrix} \\ \textsc{cont} & \ldots \begin{bmatrix} \textsc{inst} & X \\ \textit{grammar} & \end{bmatrix} \\ \textit{loc} & \end{bmatrix} \\ \textit{word} & \end{bmatrix}$$

The part of speech of *grammar* is *noun*. In order to form a complete phrase, it requires a determiner. This is represented by giving the SPR feature the value ⟨ DET ⟩. Semantic information is listed under CONT. For details see Chapter 9.

Since we are dealing with syntactic aspects exclusively, only a subset of the used features is relevant: valence information and information about part of speech and certain morphosyntactic properties that are relevant for the external distribution of a phrase is represented in a feature description under the path SYNSEM|LOC|CAT. The features that are particularly interesting here are the so-called head features. Head features are shared between a lexical head and its maximal projection. The head features are located inside CAT and are grouped together under the path HEAD. Complex hierarchical structure is also modelled with feature value pairs. The constituents of a complex linguistic object are usually represented as parts of the representation of the complete object. For instance, there is a feature HEAD-DAUGHTER the value of which is a feature structure that models a linguistic object that contains the head of a phrase. The Head Feature Principle (50) refers to this daughter and ensures that the head features of the head daughter are identical with the head features of the mother node, that is, they are identical to the head features of the complete object.

(50) *headed-phrase* ⇒
$$\begin{bmatrix} \textsc{synsem|loc|cat|head} & \boxed{1} \\ \textsc{head-dtr|synsem|loc|cat|head} & \boxed{1} \end{bmatrix}$$

Identity is represented by boxes with the same number.

Ginzburg & Sag (2000: 30) represent all daughters of a linguistic object in a list that is given as the value of the DAUGHTERS attribute. The value of the feature HEAD-DAUGHTER is identified with one of the elements of the DAUGHTERS list:

(51) a. $\begin{bmatrix} \textsc{head-dtr} & \boxed{1} \\ \textsc{dtrs} & \langle \boxed{1}\,\alpha, \beta \rangle \end{bmatrix}$

b. $\begin{bmatrix} \textsc{head-dtr} & \boxed{1} \\ \textsc{dtrs} & \langle \alpha, \boxed{1}\,\beta \rangle \end{bmatrix}$

$\alpha$ and $\beta$ are shorthands for descriptions of linguistic objects. The important point about the two descriptions in (51) is that the head daughter is identical to one of the two daughters, which is indicated by the $\boxed{1}$ in front of $\alpha$ and $\beta$, respectively. In the first feature description, the first daughter is the head and in the second description, the second daughter is the head. Because of the Head Feature Principle, the syntactic properties of the whole phrase are determined by the head daughter. That is, the syntactic properties of the head daughter correspond to the label in Chomsky's definition. This notation corresponds exactly to the one that is used by Stabler: (51a) is equivalent to (52a) and (51b) is equivalent to (52b).

(52) a. <


b. >


An alternative structuring of this basic information, discussed by Pollard & Sag (1994: Chapter 9), uses the two features HEAD-DAUGHTER and NON-HEAD-DAUGHTERS rather than HEAD-DAUGHTER and DAUGHTERS. This gives rise to feature descriptions like (53a), which corresponds directly to Chomsky's set-based representations, discussed in Section 4.6.2 and repeated here as (53b).

(53) a. $\begin{bmatrix} \textsc{head-dtr} & \alpha \\ \textsc{non-head-dtrs} & \langle \beta \rangle \end{bmatrix}$

b. $\{\, \alpha, \{\, \alpha, \beta \,\} \,\}$

The representation in (53a) does not contain information about linear precedence of $\alpha$ and $\beta$. Linear precedence of constituents is constrained by linear precedence rules, which are represented independently from constraints regarding (immediate) dominance.

The definition of Internal Merge in (43) corresponds to the Head-Filler Schema in HPSG (Pollard & Sag 1994: 164). Stabler's derivational rule deletes the subtree $t_2[-\mathrm{f}]^{>}$. HPSG is monotonic, that is, nothing is deleted in structures that are licensed by a grammar. Instead of deleting $t_2$ inside of a larger structure, structures containing an empty element (NB – not a tree) are licensed directly.[34] Both in Stabler's definition and in the

[34] See Bouma, Malouf & Sag (2001a) for a traceless analysis of extraction in HPSG and Müller (2015b: Chapter 7) and Chapter 19 of this book for a general discussion of empty elements.

HPSG schema, $t_2$ is realized as filler in the structure. In Stabler's definition of Internal Merge, the category of the head daughter is not mentioned, but Pollard & Sag (1994: 164) restrict the head daughter to be a finite verbal projection. Chomsky (2007: 17) assumes that all operations but External Merge operate on phase level. Chomsky assumes that CP and v*P are phases. If this constraint is incorporated into the definition in (43), the restrictions on the label of $t_1$ would have to be extended accordingly. In HPSG, sentences like (54) have been treated as VPs, not as CPs and hence Pollard and Sag's requirement that the head daughter in the Head Filler Schema be verbal corresponds to Chomsky's restriction.

(54) Bagels, I like.

Hence, despite minor presentational differences, we may conclude that the formalization of Internal Merge and that of the Head-Filler Schema are very similar.

An important difference between HPSG and Stabler's definition is that 'movement' is not feature driven in HPSG. This is an important advantage since feature-driven movement cannot deal with instances of so-called altruistic movement (Fanselow 2003a), that is, movement of a constituent that happens in order to make room for another constituent in a certain position (see Section 4.6.1.4).

A further difference between general $\overline{\text{X}}$ theory and Stabler's formalization of Internal Merge on the one hand and HPSG on the other is that in the latter case there is no restriction regarding the completeness (or valence 'saturation') of the filler daughter. Whether the filler daughter has to be a maximal projection (English) or not (German), follows from restrictions that are enforced locally when the trace is combined with its head. This makes it possible to analyze sentences like (55) without remnant movement.[35]

(55) Gelesen$_i$ hat$_j$ das Buch keiner $\__i$ $\__j$.
read has the book nobody

In contrast, Stabler is forced to assume an analysis like the one in (56b) (see also G. Müller (1998) for a remnant movement analysis). In a first step, *das Buch* is moved out of the VP (56a) and in a second step, the emptied VP is fronted as in (56b).

(56) a. Hat [das Buch]$_j$ [keiner [$_{\text{VP}}$ $\__j$ gelesen]].
b. [$_{\text{VP}}$ $\__j$ Gelesen]$_i$ hat [das Buch]$_j$ [keiner $\__i$].

Haider (1993: 281), De Kuthy & Meurers (2001: Section 2) and Fanselow (2002) showed that this kind of remnant movement analysis is problematic for German. The only phenomenon that Fanselow identified as requiring a remnant movement analysis is the problem of multiple fronting (see Müller (2003a) for an extensive discussion of relevant data). Müller (2005b,c, 2015b) develops an alternative analysis of these multiple frontings which uses an empty verbal head in the *Vorfeld*, but does not assume that adjuncts or arguments like *das Buch* in (56b) are extracted from the *Vorfeld* constituent. Instead of the remnant

[35] See also Müller & Ørsnes (2013b) for an analysis of object shift in Danish that can account for verb fronting without remnant movement. The analysis does not have any of the problems that remnant movement analyses have.

movement analysis, the mechanism of argument composition from Categorial Grammar (Geach 1970; Hinrichs & Nakazawa 1994) is used to ensure the proper realization of arguments in the sentence. Chomsky (2007: 20) already uses argument composition as part of his analysis of TPs and CPs. Hence both remnant movement and argument composition are assumed in recent Minimalist proposals. The HPSG alternative, however, would appear to need less theoretical apparatus and hence has to be preferred for reasons of parsimony.

Finally, it should be mentioned that all transformational accounts have problems with Across the Board extraction like (57a) and (57) in which one element corresponds to several gaps.

(57) a. Bagels, I like and Ellison hates.[36]
b. The man who$_i$ [Mary loves _$_i$] and [Sally hates _$_i$] computed my tax.

This problem was solved for GPSG by Gazdar (1981b) and the solution carries over to HPSG. The Minimalist community tried to address these problems by introducing operations like sideward movement (Nunes 2004) where constituents can be inserted into sister trees. So in the example in (57a), *Bagels* is copied from the object position of *hates* into the object position of *like* and then these two copies are related to the fronted element. Kobele criticized such solutions since they overgenerate massively and need complicated filters. What he suggests instead is the introduction of a GPSG-style SLASH mechanism into Minimalist theories (Kobele 2008).

Furthermore, movement paradoxes (Bresnan 2001: Chapter 2) can be avoided by not sharing all information between filler and gap, a solution that is not available for transformational accounts, which usually assume identity of filler and gap or – as under the Copy Theory of Movement – assume that a derivation contains multiple copies of one object only one of which is spelled out. See also Borsley (2012) for further puzzles for, and problems of, movement-based approaches.

A further difference between MG and HPSG is that the Head-Filler Schema is not the only schema for analyzing long-distance dependencies. As was noted in footnote 10 on page 149, there is dislocation to the right (extraposition) as well as fronting. Although these should certainly be analyzed as long-distance dependencies, they differ from other long-distance dependencies in various respects (see Section 13.1.5). For analyses of extraposition in the HPSG framework, see Keller (1995); Bouma (1996), Müller (1999a: Chapter 13).

Apart from the schema for long-distance dependencies, there are of course other schemata in HPSG which are not present in MG or Minimalism. These are schemata which describe constructions without heads or are necessary to capture the distributional properties of parts of constructions, which cannot be easily captured in lexical analyses (e. g. the distribution of *wh-* and relative pronouns). See Section 21.10.

Chomsky (2010) has compared a Merge-based analysis of auxiliary movement to a HPSG analysis and critiqued that the HPSG analysis uses ten schemata rather than one

[36] Pollard & Sag (1994: 205).

(Merge). Ginzburg & Sag (2000) distinguish three types of construction with moved auxiliaries: inverted sentences such as those with fronted adverbial and with *wh*-questions (58a,b), inverted exclamatives (58c) and polar interrogatives (58d):

(58) a. Under no circumstances *did she think they would do that.*
b. Whose book *are you reading*?
c. Am I tired!
d. Did Kim leave?

Fillmore (1999) captures various different usage contexts in his Construction Grammar analysis of auxiliary movement and shows that there are semantic and pragmatic differences between the various contexts. Every theory must be able to account for these. Furthermore, one does not necessarily require ten schemata. It is possible to determine this – as Categorial Grammar does – in the lexical entry for the auxiliary or on an empty head (see Chapter 21 for a more general discussion of lexical and phrasal analyses). Regardless of this, every theory has to somehow account for these ten differences. If one wishes to argue that this has nothing to do with syntax, then somehow this has to be modelled in the semantic component. This means that there is no reason to prefer one theory over another at this point.

### 4.6.5 Selection of atomic features vs. selection of complex categories

Berwick & Epstein (1995) pointed out that Minimalist theories are very similar to Categorial Grammar and I have discussed the similarities between Minimalist theories and HPSG in Müller (2013c) and in the previous subsections. However, I overlooked one crucial difference between the usual assumptions about selection in Minimalist proposals on the one hand and Categorial Grammar, Dependency Grammar, LFG, HPSG, TAG, and Construction Grammar on the other hand: what is selected in the former type of theory is a single feature, while the latter theories select for feature bundles. This seems to be a small difference but the consequences are rather severe. Stabler's definition of External Merge that was given on page 166 removes the selection feature (=f) and the corresponding feature of the selected element (f). In some publications and in the introduction in this book, the selection features are called uninterpretable features and are marked with a *u*. The uninterpretable features have to be checked and then they are removed from the linguistic object as in Stabler's definition. The fact that they have been checked is represented by striking them out. It is said that all uninterpretable features have to be checked before a syntactic object is send to the interfaces (semantics and pronunciation). If uninterpretable features are not checked, the derivation crashes. Adger (2003: Section 3.6) explicitly discusses the consequences of these assumptions: a selecting head checks a feature of the selected object. It is not possible to check features of elements that are contained in the object that a head combines with. Only features at the top-most node, the so-called root node, can be checked with external merge. The only way features inside complex objects can be checked is by means of movement. This means that a head may not combine with a partially saturated linguistic object, that is,

with a linguistic object that has an unchecked selection feature. I will discuss this design decision with reference to an example provided by Adger (2003: 95). The noun *letters* selects for a P and Ps select for an N. The analysis of (59a) corresponds to the left tree in Figure 4.23.

(59) a. letters to Peter
b. * letters to

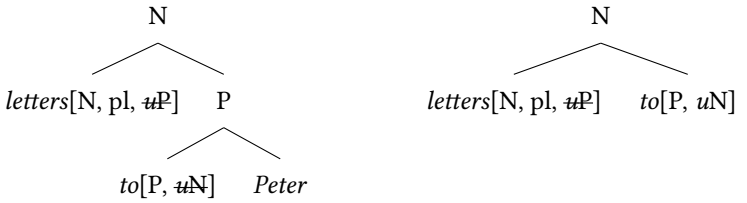



Figure 4.23: The analysis of *letters to Peter* according to Adger (2003: 95)

The string in (59b) is ruled out since the uninterpretable N feature of the preposition *to* is not checked. So this integrates the constraint that all dependent elements have to be maximal into the core mechanism. This makes it impossible to analyze examples like (60) in the most straightforward way, namely as involving a complex preposition and a noun that is lacking a determiner:

(60) vom Bus
from.the bus

In theories in which complex descriptions can be used to describe dependants, the dependent may be partly saturated. So for instance in HPSG, fused prepositions like *vom* 'from.the' can select an $\overline{\text{N}}$, which is a nominal projection lacking a specifier:

(61) N[SPR ⟨ DET ⟩]

The description in (61) is an abbreviation for an internally structured set of feature-value pairs (see Section 9.6.1). The example here is given for the illustration of the differences only, since there may be ways of accounting for such cases in a single-feature-Merge system. For instance, one could assume a DP analysis and have the complex preposition select a complete NP (something of category N with no uninterpretable features). Alternatively, one can assume that there is indeed a full PP with all the structure that is usually assumed and the fusion of preposition and determiner happens during pronunciation. The first suggestion eliminates the option of assuming an NP analysis as it was suggested by Bruening (2009) in the Minimalist framework.

Apart from this illustrative example with a fused preposition, there are other cases in which one may want to combine unsaturated linguistic objects. I already discussed coordination examples above. Another example is the verbal complex in languages like

German, Dutch, Japanese, and Korean. Of course there are analyses of these languages that do not assume a verbal complex (G. Müller 1998; Wurmbrand 2003a), but these are not without problems. Some of the problems were discussed in the previous section as well.

Summing up this brief subsection, it has to be said that the feature checking mechanism that is built into the conception of Merge is more restrictive than the selection that is used in Categorial Grammar, Lexical Functional Grammar, HPSG, Construction Grammar, and TAG. In my opinion, it is too restrictive.

### 4.6.6 Summary

In sum, one can say that the computational mechanisms of the Minimalist Program (e. g. transderivational constraints and labeling) as well as the theory of feature-driven movement are problematic and the assumption of empty functional categories is sometimes ad hoc. If one does not wish to assume that these categories are shared by all languages, then proposing two mechanisms (Merge and Move) does not represent a simplification of grammar since every single functional category which must be stipulated constitutes a complication of the entire system.

The labeling mechanism is not yet worked out in detail, does not account for the phenomena it was claimed to provide accounts for and hence should be replaced by the head/functor-based labeling that is used in Categorial Grammar and HPSG.

## 4.7 Summary and classification

This section is similar to Section 3.6. I first comment on language acquisition and then on formalization.

### 4.7.1 Explaining language acquisition

Chomsky (2008: 135) counts theories in the MP as Principle and Parameter analyses and identifies MP parameters as being in the lexicon. Also, see Hornstein (2013: 396). UG is defined as possibly containing non-language-specific components, which are genetically determined (Chomsky 2007: 7). UG consists of unbounded Merge and the condition that expressions derived by a grammar must fulfill the restrictions imposed by the phonological and conceptual-intentional interfaces. In addition, a specific repertoire of features is assumed to be part of UG (Chomsky 2007: 6–7). The exact nature of these features has not been explained in detail and, as a result, the power of UG is somewhat vague. However, there is a fortunate convergence between various linguistic camps as Chomsky does not assume that the swathes of functional projections which we encountered in Section 4.6.1 also form part of UG (however, authors like Cinque & Rizzi (2010) do assume that a hierarchy of functional projections is part of UG). Since there are still parameters, the same arguments used against GB approaches to language acquisition that were mentioned in Section 3.6.1 are still relevant for theories of language acquisition

in the Minimalist Program. See Chapter 16 for an in-depth discussion of approaches to language acquisition and the Principles and Parameters model as well as input-based approaches.

Chomsky's main goal in the Minimalist Program is to simplify the theoretical assumptions regarding formal properties of language and the computational mechanisms that are used so much as to make it plausible that they or relevant parts of them are part of our genetic endowment. But if we recapitulate what was assumed in this chapter, it is difficult to believe that Minimalist theories achieve this goal. To derive a simple sentence with an intransitive verb, one needs several empty heads and movements. Features can be strong or weak, Agree operates nonlocally in trees across several phrase boundaries. And in order to make correct predictions, it has to be made sure that Agree can only see the closest possible element (13)–(14). This is a huge machinery in comparison to a Categorial Grammar that just combines adjacent things. Categorial Grammars can be acquired from input (see Section 13.8.3), while it is really hard to imagine how the fact that there are features that trigger movement when they are strong, but do not trigger it when they are weak, should be acquired from data alone.

### 4.7.2 Formalization

Section 3.6.2 commented on the lack of formalization in transformational grammar up until the 1990s. The general attitude towards formalization did not change in the minimalist era and hence there are very few formalizations and implementations of Minimalist theories.

Stabler (2001) shows how it is possible to formalize and implement Kayne's theory of remnant movement. In Stabler's implementation[37], there are no transderivational constraints, no numerations[38], he does not assume Agree (see Fong 2014: 132) etc. The fol-

[37] His system can be downloaded from his website: http://www.linguistics.ucla.edu/people/stabler/coding.html. 31.03.2010.

[38] There is a numeration lexicon in Veenstra (1998: Chapter 9). This lexicon consists of a set of numerations, which contain functional heads, which can be used in sentences of a certain kind. For example, Veenstra assumes numerations for sentences with bivalent verbs and subjects in initial position, for embeded sentences with mono-valent verbs, for *wh*-questions with mono-valent verbs and for polar interrogatives with mono-valent verbs. An element from this set of numerations corresponds to a particular configuration and a phrasal construction in the spirit of Construction Grammar. Veenstra's analysis is not a formalization of the concept of the numeration that one finds in Minimalist works. Normally, it is assumed that a numeration contains all the lexical entries which are needed for the derivation of a sentence. As (i) shows, complex sentences can consist of combinations of sentences with various different sentence types:

(i) Der Mann, der behauptet hat, dass Maria gelacht hat, steht neben der Palme, die im letzten
the man who claimed has that Maria laughed has stands next.to the palm.tree which in last
Jahr gepflanzt wurde.
year planted was
'The man who claimed that Maria laughed is standing next to the palm tree that was planted last year.'

In (i), there are two relative clauses with verbs of differing valence, an embedded sentence with a mono-valent verb and the matrix clause. Under a traditional understanding of numerations, Veenstra would have to assume an infinite numeration lexicon containing all possible combinations of sentence types.

lowing is also true of Stabler's implementation of Minimalist Grammars and GB systems: there are no large grammars. Stabler's grammars are small, meant as a proof of concept and purely syntactic. There is no morphology[39], no treatment of multiple agreement (Stabler 2011: Section 27.4.3) and above all no semantics. PF and LF processes are not modelled.[40] The grammars and the computational system developed by Sandiway Fong are of similar size and faithfulness to the theory (Fong & Ginsburg 2012; Fong 2014): the grammar fragments are small, encode syntactic aspects such as labeling directly in the phrase structure (Fong & Ginsburg 2012: Section 4) and therefore fall behind $\overline{\text{X}}$ theory. Furthermore, they do not contain any morphology. Spell-Out is not implemented, so in the end it is not possible to parse or generate any utterances.[41] The benchmark here has been set by implementations of grammars in constraint-based theories; for example, the HPSG grammars of German, English and Japanese that were developed in the 90s as part of Verb*mobil* (Wahlster 2000) for the analysis of spoken language or the LFG or CCG systems with large coverage. These grammars can analyze up to 83 % of utterances in spoken language (for Verb*mobil* from the domains of appointment scheduling and trip planning) or written language. Linguistic knowledge is used to generate and analyze linguistic structures. In one direction, one arrives at a semantic representation of a string of words and in the other one can create a string of words from a given semantic representation. A morphological analysis is indispensable for analyzing naturally occurring data from languages with elaborated morphological marking systems. In the remainder of this book, the grammars and computational systems developed in other theories will be discussed at the beginning of the respective chapters.

The reason for the lack of larger fragments inside of GB/MP could have to do with the fact that the basic assumptions of Minimalist community change relatively quickly:

> In Minimalism, the triggering head is often called a *probe*, the moving element is called a *goal*, and there are various proposals about the relations among the features that trigger syntactic effects. Chomsky (1995b: p. 229) begins with the assumption

[39] The test sentences have the form as in (i).

(i) a. the king will -s eat
b. the king have -s eat -en
c. the king be -s eat -ing
d. the king -s will -s have been eat -ing the pie

[40] See Sauerland & Elbourne (2002) for suggestions of PF and LF-movement and the deletion of parts of copies (p. 285). The implementation of this would be far from trivial.

[41] The claim by Berwick, Pietroski, Yankama & Chomsky (2011: 1221) in reference to Fong's work is just plain wrong: *But since we have sometimes adverted to computational considerations, as with the ability to "check" features of a head/label, this raises a legitimate concern about whether our framework is computationally realizable. So it is worth noting that the copy conception of movement, along with the locally oriented "search and labeling" procedure described above, can be implemented computationally as an efficient parser; see Fong, 2011, for details.* If one has a piece of software which cannot parse a single sentence, then one cannot claim that it is efficient since one does not know whether the missing parts of the program could make it extremely inefficient. Furthermore, one cannot compare the software to other programs. As has already been discussed, labeling is not carried out by Fong as was described in Chomsky's work but instead he uses a phrase structure grammar of the kind described in Chapter 2.

> that features represent requirements which are checked and deleted when the requirement is met. The first assumption is modified almost immediately so that only a proper subset of the features, namely the 'formal', 'uninterpretable' features are deleted by checking operations in a successful derivation (Collins, 1997; Chomsky 1995b: §4.5). Another idea is that certain features, in particular the features of certain functional categories, may be initially unvalued, becoming valued by entering into appropriate structural configurations with other elements (Chomsky 2008; Hiraiwa, 2005). And some recent work adopts the view that features are never deleted (Chomsky 2007: p. 11). These issues remain unsolved. (Stabler 2010: 397)

In order to fully develop a grammar fragment, one needs at least three years (compare the time span between the publication of *Barriers* (1986) and Stabler's implementation (1992)). Particularly large grammars require the knowledge of several researchers working in international cooperation over the space of years or even decades. This process is disrupted if fundamental assumptions are repeatedly changed at short intervals.

## Further reading

This chapter heavily draws from Adger (2003). Other textbooks on Minimalism are Radford (1997), Grewendorf (2002), and Hornstein, Nunes & Grohmann (2005).

Kuhn (2007) offers a comparison of modern derivational analyses with constraint-based LFG and HPSG approaches. Borsley (2012) contrasts analyses of long-distance dependencies in HPSG with movement-based analyses as in GB/Minimalism. Borsley discusses four types of data which are problematic for movement-based approaches: extraction without fillers, extraction with multiple gaps, extractions where fillers and gaps do not match and extraction without gaps.

The discussion of labeling, abandonment of $\overline{\text{X}}$ theory and a comparison between Stabler's Minimalist Grammars and HPSG from Sections 4.6.2–4.6.4 can be found in Müller (2013c).

*Intonational Phrasing, Discontinuity, and the Scope of Negation* by Błaszczak & Gärtner (2005) is recommended for the more advanced reader. The authors compare analyses of negated quantifiers with wide scope in the framework of Minimalism (following Kayne) as well as Categorial Grammar (following Steedman).

Sternefeld (2006) is a good, detailed introduction to syntax (839 pages) which develops a Transformational Grammar analysis of German which (modulo transformations) almost matches what is assumed in HPSG (feature descriptions for arguments ordered in a valence list according to a hierarchy). Sternefeld's structures are minimal since he does not assume any functional projections if they cannot be motivated for the language under discussion. Sternefeld is critical regarding certain aspects which some other analyses take for granted. Sternefeld views his book explicitly as a textbook from which one can learn how to argue coherently when creating theories. For this reason, this book is not just recommended for students and PhD students.

Sternefeld & Richter (2012) discuss the situation in theoretical linguistics with particular focus on the theories described in this and the previous chapter. I can certainly understand the frustration of the authors with regard to the vagueness of analyses, argumentation style, empirical base of research, rhetorical clichés, immunization attempts and general respect for scientific standards: a current example of this is the article *Problems of Projection* by Chomsky (2013).[42] I, however, do not share the general, pessimistic tone of this article. In my opinion, the patient's condition is critical, but he is not dead yet. As a reviewer of the Sternefeld and Richter paper pointed out, the situation in linguistics has changed so much that now having a dissertation from MIT does not necessarily guarantee you a position (footnote 16) later on. One could view a reorientation of certain scientists with regard to certain empirical questions, adequate handling of data (Fanselow: 2004b; 2009: 137) and improved communication between theoretical camps as a way out of this crisis.

Since the 90s, it is possible to identify an increased empirical focus (especially in Germany), which manifests itself, for example, in the work of linguistic Collaborative Research Centers (SFBs) or the yearly *Linguistic Evidence* conference. As noted by the reviewer cited above, in the future it will not be enough to focus on Chomsky's problems in determining the syntactic categories in sentences such as *He left* (see Section 4.6.2). Linguistic dissertations will have to have an empirical section, which shows that the author actually understands something about language. Furthermore, dissertations, and of course other publications, should give an indication that the author has not just considered theories from a particular framework but is also aware of the broad range of relevant descriptive and theoretical literature.

As I have shown in Section 4.6.4 and in Müller (2013c) and will also show in the following chapters and the discussion chapters in particular, there are most certainly similarities between the various analyses on the market and they do converge in certain respects. The way of getting out of the current crisis lies with the empirically-grounded and theoretically broad education and training of following generations.

For short: both teachers and students should read the medical record by Sternefeld and Richter. I implore the students not to abandon their studies straight after reading it, but rather to postpone this decision at least until after they have read the remaining chapters of this book.

[42] Vagueness: in this article, *perhaps* occurs 19 times, *may* 17 as well as various *if*s. Consistency: the assumptions made are inconsistent. See footnote 16 on page 158 of this book. Argumentation style: the term specifier is abolished and it is claimed that the problems associated with this term can no longer be formulated. Therefore, they are now not of this world. See footnote 26 on page 162 of this book. Immunization: Chomsky writes the following regarding the Empty Category Principle: *apparent exceptions do not call for abandoning the generalization as far as it reaches, but for seeking deeper reasons to explain where and why it holds* p. 9. This claim is most certainly correct, but one wonders how much evidence one needs in a specific case in order to disregard a given analysis. In particular regarding the essay *Problems of Projection*, one has to wonder why this essay was even published only five years after *On phases*. The evidence against the original approach is overwhelming and several points are taken up by Chomsky (2013) himself. If Chomsky were to apply his own standards (for a quote of his from 1957, see page 8) as well as general scientific methods (Occam's Razor), the consequence would surely be a return to head-based analyses of labeling.

For detailed comments on this essay, see Sections 4.6.2 and 4.6.3.